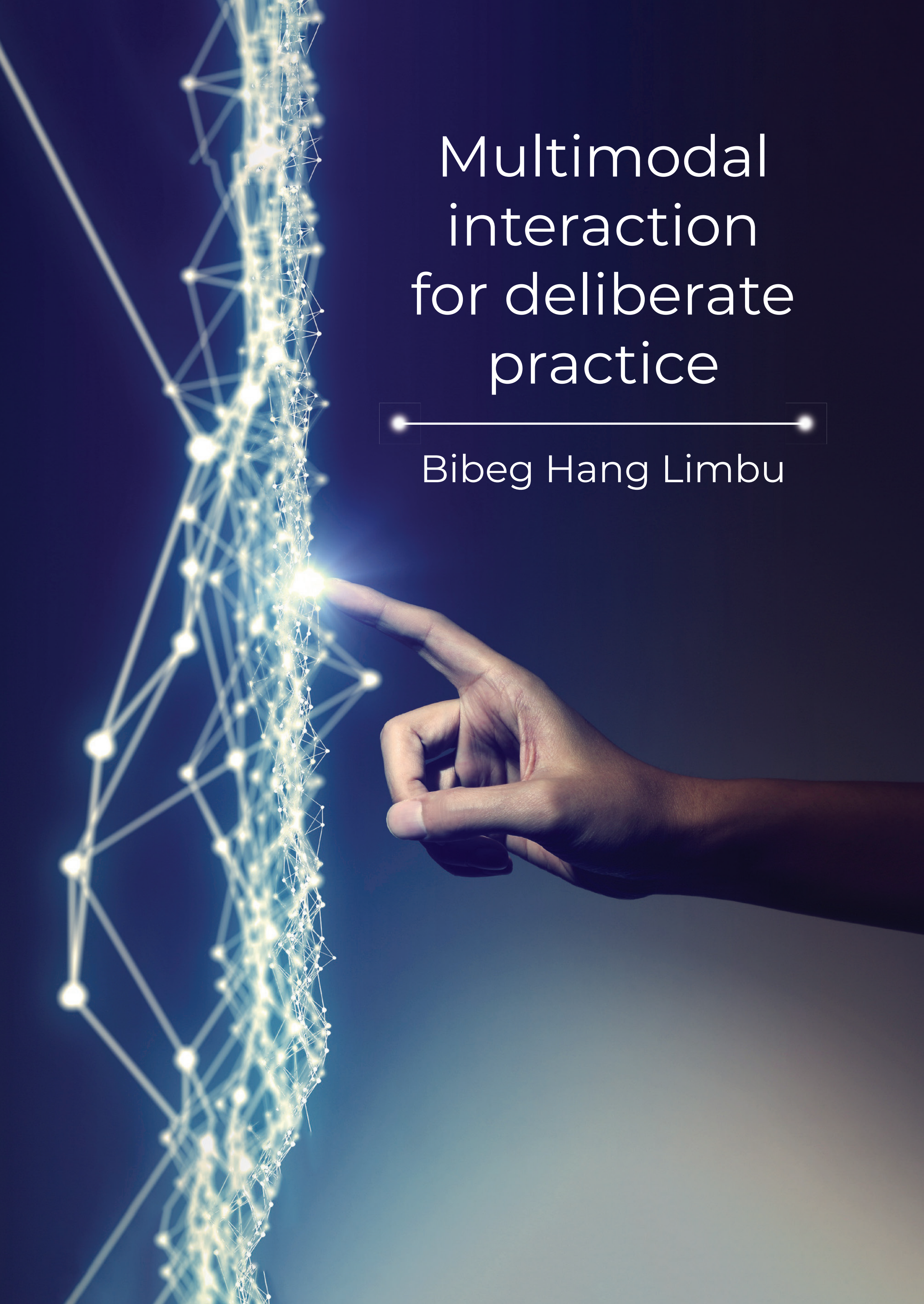
Multimodal interaction for deliberate practice



Multimodal interaction for deliberate practice

Bibeg Hang Limbu

Multimodal interaction for deliberate practice

Training complex skills with augmented reality

The research reported in this thesis was carried out at the Open Universiteit in the Netherlands at the Faculty of Education, formerly known as Welten Institute – Research Centre for Learning, Teaching and Technology,

and under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems.

SIKS Dissertation Series No. 2020-28

Printed by ProefschriftMaken
Cover design: ProefschriftMaken
Typeset in LaTeX

ISBN/EAN: 978-94-6380-963-4

Multimodal interaction for deliberate practice

Training complex skills with augmented reality

Proefschrift

ter verkrijging van de graad van doctor
aan de Open Universiteit
op gezag van de rector magnificus
prof. dr. Th.J. Bastiaens
ten overstaan van een door het
College voor promoties ingestelde commissie
in het openbaar te verdedigen

op vrijdag 6 november 2020 te Heerlen
om 13:30 uur precies

door

Bibeg Hang Limbu
geboren op 28 december 1988, Hongkong

Promotor

Prof. dr. M.M. Specht
Open Universiteit / TU Delft

Co-Promotorers

Prof. dr. R. Klemke
Open Universiteit / TH Köln

Prof. dr. H. Jarodzka
Open Universiteit

Leden beoordelingscommissie

Prof. dr. E. Prasolova-Førland
Norwegian University (of Science and Technology)
Prof. dr. A. Weinberger
University of Saarland
Prof. dr. H.J. Drachsler
Open Universiteit / DIPF - Leibniz Institute for Research and Information in Education
Prof. dr. ir. C.J. Kreijns
Open Universiteit

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Overview

Increasing technological and societal innovations are generating higher demands on students to master many complex skills needed in the industry (Frerejean et al., 2019). The set of skills required to work in an industry is continually growing broader and more complex, as the economy is transforming digitally. Simple skills are being automated and the demand for complex skills is rising (Garbi, 2020). 54% of companies across the globe report talent shortages (ManpowerGroup, 2020). On the other hand, the higher educational institutions are burdened with an increasing number of students they have to train with limited resources. Students are unable to master all the complex skills during their formal education. Additionally, more than ever, mastery of complex skills happen in informal contexts, such as through internship or self-practice. Moreover, the proficiency levels in various domains, in medicine for example, have continuously risen over the past years. This phenomena is creating an urgent need to support students, and lifelong learners, in developing their complex skills both in formal and informal learning contexts.

Complex skills are difficult to learn and are a comprised set of constituent skills which require conscious processing and an estimated five hundred hours to acquire (Ackermans et al., 2016). All the constituent skills of a complex skill contribute towards achieving the main objective of the complex skill. Complex skills are valuable human resources as it also has been argued that they can not be fully automated by a machine (Garbi, 2020). However, Van Merriënboer (1997) states that not all constituent sub-skills of a complex skill are performed in the same manner. Some constituent sub-skills of a complex skill involve conscious processing and therefore require a large attention span during practice (van Merriënboer and Kester, 2014) while some of them need to be automated by repeated practice. For example, surgery as a complex skill requires, among other, the skills of stitching, anatomy proficiency and dexterity. Without one of these sub-skills, the surgeon cannot perform surgery successfully. Some of these skills such as dexterity can be automated by repeated practice while others such as anatomy proficiency requires conscious processing. Because some sub-skills cannot be automated, but require conscious processing, the overall complex skill can never be automated by definition. But it can be more or less efficiently executed. Constituent skills which require automatic processing require a different learning process as compared to skills which require conscious processing. Consequently, complex skills require a lot of time and effort to master, which can be improved with a proper use of instructional design.

The Four Components Instructional Design (4C/ID)(Van Merriënboer et al.,

2002) aims to make learning of complex skills more efficient. The basic assumption of the 4C/ID is that complex skills learning can be described in terms of four components, namely

Learning tasks : whole task experiences based on authentic tasks,

Supportive information : information that is supportive to the learning and performance of problem solving and reasoning aspects of learning tasks,

Procedural information : information that is prerequisite to the learning and performance of routine aspects of learning tasks,

Part-task practice : additional exercises for routine aspects of learning tasks for which a very high level of automaticity is required after the instruction van Merriënboer and Kester (2014).

Van Merriënboer and Kirschner (2017) provide a practicable version of the 4C/ID model for mentors and instructional designers to make it more applicable to educational practice. In addition to simplifying the instructional design of complex skills training, the 4C/ID model also has a close resemblance with underlying principles of deliberate practice (Neelen and Kirschner, 2016) which is vital for effective and efficient mastery of complex skills.

Lee et al. (2018) state that simply practising a skill does not guarantee improvement of the skill and that practice must be deliberate. Deliberate practice involves conscious executions aimed at improving a particular skill (Ericsson et al., 1993). Gladwell (2008) in his popular book "Outliers", proposed that a minimum of 10,000 hours of practice was necessary to become an expert but this has been found to be only partially true in other studies (Ericsson and Harwell, 2019). Students need to practice *deliberately* to improve their skill optimally (Ericsson et al., 2018). In his updated list of 250+ factors that influence student achievement, Hattie (2017) included deliberate practice as a factor which is beyond the premises of educational institutions but with a high effect size (d = 0.79) on training outcomes. Thus, it is evident that students must practice deliberately to master skills efficiently and effectively.

However, deliberate practice develops over a prolonged period when a motivated individual challenges him/herself with a desire to improve their skill. Motivation of students, among other factors such as optimised cognitive load, is vital to practice deliberately (McGaghie et al., 2010). At the same time, it is difficult for novice students to practice deliberately as it is cognitively demanding to be conscious of their own performance (Rikers et al., 2004; Ericsson et al., 2007). Therefore, Ericsson et al. (2018) stressed the importance of an expert mentor for deliberate practice, stating that students do not engage in deliberate practice spontaneously. Traditionally, and in many ways even today, complex skills are still learnt by means of an apprenticeship where students learn from a mentor by observation, imitation and modelling (Collins et al., 1988). A key aspect of such apprenticeship, is the deep involvement of the mentor in students' repeated authentic practice simulating a one to one settings. By doing so, the mentor plays a crucial role in enabling deliberate practice in students so that they achieve mastery of the skill. This is inline with Ericsson and Harwell (2019), who emphasises individualised training by a well-qualified mentor as a key aspect of deliberate practice. However, mentors are scarce and costly. With increasing number of students and course contents, mentors cannot provide enough attention to individual students. Moreover, the deliberate practice of complex skills must be done in authentic settings (Neelen and Kirschner, 2018) which requires a large amount of physical resources. It is difficult to provide the required physical resources for deliberate practice to every individual student which negatively affects how often students practice deliberately and thus, results in poor achievements. This shortcoming can potentially be addressed by using educational technology to support deliberate practice (Han et al., 2015).

Sarfo and Elen (2006) assessed educational learning environments developed with 4C/ID specifications and positively indicated that the 4C/ID model promoted deliberate practice. Evidence about the effectiveness of such systems for promoting deliberate practice in training contexts has also been documented by Merriënboer and Paas (2003) and Merrill (2002). Moreover, van Merriënboer and Kester (2014) have also discussed the use of the 4C/ID model to design educational learning environments in which instruction is controlled by the system, the learner/mentor, or both. To conclude, educational technology can support deliberate with 4C/ID model based instructions, however, to train complex skills, authentic practice is required. One way to meet these requirements is by using augmented reality (AR) as the medium for educational learning environments (Bacca et al., 2014).

AR superimposes computer-generated layers of digital information on top of temporal and physical space of the user. However, this conventional definition of AR heavily emphasises on the virtual visual information being augmented on top of the physical environment and the richness of AR as a truly immersive medium is lost in this simplification of the concept (Papagiannis, 2017). For AR to be a truly immersive medium, it must put the user at the centre of the interaction and therefore, should embrace a multimodal approach of interaction to simulate and streamline naturalistic interaction between the physical and the virtual environment (Schraffenberger and van der Heide, 2016). To achieve this, AR needs to be context aware of both the physical and the virtual environment and the interaction between them, and therefore, must incorporate sensors, which have the ability to monitor and measure physical properties, as its fundamental unit. This provides various affordances for educational use of AR to support authentic practice (Bacca et al., 2014). For example, AR can replace physical objects with virtual 3D models which allow students to have an repeated authentic practice. In addition, AR also supports intuitive computer human interactions such as hand gestures which makes practice with virtual 3D models more authentic. In this way, AR can potentially address the difficulty of providing physical resources required for deliberate practice in authentic settings while reducing associated costs and risks (Bacca et al., 2014).

Additionally, Schneider et al. (2015) in their review of multimodal applications in the domain of learning and training have outlined the capabilities of sensors in providing feedback. Multimodal applications such as AR along with various sensors provide a rich multimodal, multisensory medium for students to learn in an authentic settings. Sensors, an integral part of AR (*AR and sensors together will be just stated as AR from here on-wards*), have the capability to unobtrusively monitor physical properties. AR can, therefore, be used to provide immediate feedback, a key aspect of deliberate practice, in an unobtrusive manner to students without exerting excessive mental effort during practice (Gordienko et al., 2017). AR can take over rudimentary tasks such as providing immediate feedback in deliberate practice repetitions so that mentors can use his/her precious time on more important matters such as the planning the next practice session. Planning practice according to students needs and competencies consumes significant effort and time of mentors but offer more individualisation opportunities. In addition, AR can also record performance to create expert models that can be used to provide guidance and feedback to students. For example, Jarodzka et al. (2013) used eye tracking sensors to record expert's information screening performance and generated an expert model for training. Using AR to create expert models can further reduce the time required by the experts to orchestrate practice as it abolishes the need for repeated demonstrations. It also increases the amount of time students can use the expert model to learn from it. In addition, expert models generated make remote or offline training possible as students can continue to practice without the presence of the mentor while receiving expert-like feedback and guidance. It can also record students performance so that mentors can use the data to better orchestrate practice. Thus, AR with 4C/ID based instruction has the potential to support deliberate practice of complex skills by providing authentic practice opportunities and various affordances to complement both mentors and students.

It is evident that AR can support deliberate practice of complex skills but to the best of our knowledge, there were no previous works which sought to systematically explore and evaluate its potential. Therefore, we aim to address this gap by means of design based approach which entails developing and testing AR applications to support deliberate practice of complex skills. However, inappropriate design of AR applications can affect students' progress negatively, slowing down learning or may even lead to learning of improper techniques (Barnes, 1987). AR applications must be designed with careful considerations of the affordances and theoretical implications, such that deliberate practice can be fostered. Without a proper design approach, this can be a complicated task. In search of the solution to this problem, this thesis defines the context of the research as follows.

> "How can AR applications be designed for deliberate practice of complex skills?"

To help answer this question, we derived four sub-ordinate questions that are addressed in the upcoming chapters.

> **RQ1** Which design patterns can be used in AR to train different types of skills?
>
> **RQ2** How can design patterns be systematically implemented in AR to support deliberate practice?
>
> **RQ3** How can expert performance be modelled and evaluated?
>
> **RQ4** How can feedback be designed without imposing high mental effort on students?

1.1 Outline of the Research

AR is by no means a new technology and has often been used for training various types of skills. In order to get an overview of the types of AR instructional design patterns used in training in the recent literature, chapter 2, **"Literature study"**, describes a literature review exploring various AR prototypes. The prototypes were only included in the review if a human mentor was involved in the training process. Abstract design patterns were extracted from prototypes which support training by recording mentors performance data and then, using the recorded data to train students. After getting an overview of the types of AR instructional design patterns, the chapter further seeks to explore the potential of AR for supporting deliberate practice. Therefore, we further analysed the AR instructional design patterns based on the Four Component Instructional Design (4C/ID) model (Van Merriënboer, 1997). The findings of the review show that AR can potentially support training of complex skills with instructions based on 4C/ID model, eventually supporting deliberate practice in authentic settings.

In chapter 3, **"The ID4AR framework"**, we develop an abstract conceptual framework for designing AR applications that support training of complex skills. Using our findings from the literature review, we extend the 4C/ID model into Instructional Design for Augmented Reality (ID4AR) model by encapsulating the 4C/ID model with AR based design patterns. This chapter focuses on the methodology for operationalisation of the framework. To assist the designers in the operationalisation, we create step-by-step guidelines. These guideline are meant to help instructional designers from the first step of task analysis which is required to identify the key aspects of the complex skill, to the use of the developed application in the intended manner.

The ID4AR framework provides a systematic approach to design AR applications for deliberate practice of complex skills. Using the ID4AR framework, WEKIT.One prototype was designed and developed in the context of WEKIT project for the three WEKIT domains (astronaut training, medical training, and aerospace engineering). The prototype was built to cater to the three WEKIT domains by implementing the instructional design patterns which were identified by conducting task analysis with the mentors in the three domains. The assumption of the ID4AR framework is that the instructional design patterns implemented in WEKIT.One meets the didactic requirements of their respective components of the 4C/ID model to which they are assigned. Chapter 4, **"WEKIT.One: User study"**, reports on the user study which evaluates if the WEKIT.One prototype meets this assumption. Mentors other than those that took part in the task analysis and students participated in the study. The findings suggest that both mentors and students agreed that the prototype met the ID4AR frameworks assumption. Moreover, the study also found the usability of the prototype to be acceptable.

Chapter 5, **"WEKIT.One: Expert model evaluation"**, reports on the evaluation of the expert model captured with the WEKIT.One prototype. In layman's terms, it intends to evaluate the expert model's usefulness for training. This study was also conducted in the three domains of the WEKIT project but was conducted at the later stage of the project, with a finalised prototype, after incorporating findings from the study reported in chapter 4. To conduct this study, an expert from each of the three domains used the WEKIT.One prototype to record an expert model for their respective domains. These mentors were given as much time as needed to create the expert model. However, once the expert model was finalised no further post processing was done to the learning material. This expert model was then evaluated by other mentors from the same domains to test its suitability for training.

The mentors agreed that the expert model created using the prototype was suitable for training and by that, also establishes the prototype's ability to record an expert model in different domains. Similarly, students from the three domains participated in the study in control and treatment groups. The students in the treatment practised the complex skill using the expert model for a short amount of time, while the control group was given printed manuals. The findings suggested no difference in the score of the paper-based posttest evaluating their knowledge of the procedure between the two groups.

As a next step, the thesis explores the applicability of the ID4AR framework in a new domain and the consequences of feedback given based on the framework for cognitive load of learners. Chapter 6, **"Calligraphy trainer: Assessing mental effort"**, reports on the design and development of a new prototype called "Calligraphy trainer" for teaching calligraphy to novices, using the ID4AR framework. Calligraphy is a complex perceptual-motor skill that requires many hours of practice to master (Feder and Majnemer, 2007) and thus the effects of deliberate practice can be more pronounced. The study reported in this chapter assesses the amount of mental effort induced by the immediate feedback given by the calligraphy trainer on the novices as novices find it difficult to practice deliberately due to the higher requirement of cognitive effort (Ericsson et al., 2007). The study also evaluates the usability of the prototype to ensure that no unnecessary mental effort is added as a result of poor usability or overload in multimodal feedback. The findings of this study concluded that the mental effort required to process the immediate feedback using the calligraphy trainer was not significantly different than practising calligraphy without the immediate feedback in a normal tablet application. The usability of the prototype was also found to be acceptable.

The **General discussion** section reviews the general findings of the dissertation and its limitations. It provides a summary of the findings of the previous chapters and their significance in the larger context and also suggest future paths for the research and discuss how some of the findings can be generalised into practice. The chapter concludes by expressing the author's insights and opinions regarding the use of multimodal applications for training complex skills taking the current technological progress into consideration.

1.2 Wearable Experience for Knowledge Intensive Training

The research reported in this thesis was partly funded by the WEKIT project (Wearable Experience for Knowledge Intensive Training). WEKIT was a European research and innovation project funded by the Horizon 2020 programme under grant agreement no 687669 to develop and test within three years a novel way of industrial training. WEKIT aimed to build a ground-breaking industrial-strength training application and a unique methodology to capture expert experience and share it with students in the process of enabling immersive, in-situ, and intuitive learning. The goal was to design and developed cutting edge solutions for training complex skills in the domain of Medicine, Aerospace and Astronaut training. The WEKIT.One prototype used in the study reported in chapter 3 and 4 was developed in the context of this project. Similarly the studies reported in chapter 3 and 4 were also conducted in the three WEKIT domains. Consequently, the studies had to also be limited within the scope of the project. Some of the results of this thesis contributed to the WEKIT deliverables.

Chapter 2

Literature study

What is a pirates favourite tech? AR...!!

This chapter explores the state-of-the-art of augmented reality and sensor-based design patterns for training complex skills. It systematically reviews 78 studies utilising augmented reality prototypes that made use of human mentors for training. To get an overview of the potential of how augmented reality can be used for training, instructional design patterns were extracted from these studies. Further analysis was done to understand how the instructional design patterns can support deliberate practice of complex skills. To do so, the instructional design patterns were then, analysed according to the four components instructional design pattern. The findings outline a methodological approach to use sensors and augmented reality for designing augmented reality environments for training complex skills.

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

and is also based on: Limbu, B., Fominykh, M., Klemke, R., Specht, M., and Wild, F. (2018a). *Supporting Training of Expertise with Wearable Technologies: The WEKIT Reference Framework*, pages 157–175. Springer Singapore, Singapore

2.1 Introduction

Bloom and Sosniak (1985) investigated the childhoods of 120 elite performers and observed that all of them had practised intensively and trained with devoted mentors. Ericsson et al. (2007) have also emphasised the importance of experts as mentors for apprentices. While experts seem imperative to train apprentices to achieve superior performance, they are not free of shortcomings. Experts tend to underestimate how difficult a task can be for apprentices (Hinds, 1999). Experts are also often unaware of all the knowledge behind their superior performance (Patterson et al., 2010) and thus omit information that apprentices may find valuable (Hinds et al., 2001). This may cause apprentices to find it difficult to learn from experts. Moreover, difficulty also lies in the shortage of experts to train apprentices in one-to-one settings (Bloom and Sosniak, 1985).

According to Ericsson et al. (2018), an expert is someone who consistently demonstrates superior performance in a representative set of tasks for the domain of his/her expertise. We use the term "expert performance" in the context of this paper to describe the performance of an expert. Representative tasks are structured and managed drills which elicit the same set of knowledge and skills that would be necessary in the corresponding real-world tasks. Representative tasks allow for the quantification of expert performance, which can then be analysed to assess its mediating mechanisms. This enables us to measure and capture performance, which can be used in various ways for training apprentices.

To train apprentices with captured expert performance data, we first aim to capture a rich representation of the expert performance using sensor technology. The reason to focus on sensors is that they have the capability to unobtrusively measure observable properties, which is ideal for capturing expert performance in representative tasks. Bower and Sturman (2015) have elaborated on the educational affordances of sensors and their potentials for training. A sensor is commonly defined as a device that detects or measures a physical property and records, indicates, or otherwise responds to it. However, expertise is a much broader term which encompasses tacit skills that cannot be captured directly by sensors. Regardless, sensors have already been successfully used to train apprentices based on expert performance data (e.g. (Jarodzka et al., 2013); (Schneider et al., 2017)). These studies have relied on the manual encoding of recorded data between the capture and training phases. With this study, we aim to outline instructional designs which do not require manual encoding of sensor data. Moreover, technology such as augmented reality (AR) offers important affordances such as contextual awareness, which allows for the meaningful recording and use of sensor data by integrating the sensor data into the interaction (Guest et al., 2017), potentially reducing the manual phase between capture and training.

AR overlays the real world with virtual content to create an immersive platform which places the trainee in a real-world context, engaging all his/her senses (Bacca et al., 2014). Modern AR systems can communicate with various sensors in real time, which can offer a broad range of training affordances. We term such a combination of sensors and AR as "sensor-based AR" in the context of this paper. Sensor-based AR training environments can support apprentices by providing personalised guidance and feedback when experts are not available. Sensor-based AR environments also provide a rich multimodal and multi-sensory medium for apprentices to learn efficiently from expert performance data. While (Bacca et al., 2014) suggested, based on their review, that sensor-based AR posits a rich educational potential for the personalization of training, Bower and Sturman (2015) have stressed the risk of putting technology before pedagogy. This is particularly true for emerging technologies, such as sensor-based AR, that provide a huge range of affordances which may or may not be beneficial for training and education. Therefore, to ensure we are guided by proper pedagogy, we structured our exploration of sensor-based AR with a pedagogical framework known as the Four Components Instructional Design (4C/ID) model. The 4C/ID model is a non-linear and systematic processing model for designing a complex learning environment (Van Merriënboer et al., 2002). The basic assumption of the 4C/ID model is that all complex learning can be represented in a combination of four components described by the model, namely:

2.1.1 Learning task:

Learning tasks are authentic, whole-task experiences that are provided to apprentices to promote schema construction for non-recurring aspects of the task. Construction of mental models by the apprentice can be facilitated by observing or imitating the expert. For example, by first demonstrating examples of how a particular concept is used before tackling the problem, apprentices are provided with the opportunity to construct relationships between chunks of knowledge gained during observation and imitation.

2.1.2 Supportive information:

Supportive information supports apprentices to deeply process new information for the construction of mental models. It provides domain-related information, such as approaches to solving problems rather than procedural information, to support the learning and the performance of nonrecurring aspects of the task.

2.1.3 Just in time information:

Just in time information is the prerequisite procedural information to the learning and performance of recurring aspects of the learning task provided to the apprentice precisely when required. Procedural information is immediately diminished for the subsequent task to facilitate automation of the task.

2.1.4 Part-task practise

The last component of the 4C/ID model is the part-task practice, which recognises that some parts of the task are automatic and recurrent. To automate recurring aspects of the task, it is necessary that apprentices practise these recurrent tasks repeatedly. Part-task practice items are provided to apprentices to promote rule automation for recurring aspects of the whole complex skill by means of "strengthening", in which cognitive rules accumulate higher strength upon repeated successful executions.

We selected the 4C/ID model as an appropriate framework to guide this study due to its potential to support training in complex skills, as well as due to its close resemblance to the underlying principles of deliberate practice (Neelen and

Kirschner, 2016). McDaniel et al. (1988), as well as Ericsson et al. (2018), found that only individuals who indulged in deliberate practice attained superior performance. Thus, practice should be deliberate, that is, aimed at improving that particular skill by reflecting on previous performance and collecting new experience. However, it is difficult to perform or maintain deliberate practice because it is cognitively demanding for the apprentice to be conscious of his/her own performance (Rikers et al., 2004). Expert mentors are crucial to support deliberate practice in apprentices (Ericsson et al., 2007). Deliberate practice requires one-to-one settings where an expert continuously provides guidance and feedback to the apprentice (Carey, 2014).

In conclusion, this literature study aims to investigate the potential of sensor-based AR to support deliberate practice in apprentices with the help of captured expert performance data. Therefore, we examine patterns in the literature that exploit the affordances of sensors for enabling the capture of expert performance. We also explore the use of sensor-based AR for training apprentices using expert performance data. As such we examine the following questions:

- How can sensors be used to capture expert performance?
- How can sensor-based AR utilize expert performance data to facilitate 4C/ID model-based training programs?

2.2 Methods

In order to explore how expert performance has been captured with sensors, we selected studies that either explicitly used sensors to capture expert performance or described systems which used manually-created expert models through task analysis. Similarly, we also chose studies describing systems that have the potential to use experts as a reference for training. To obtain relevant articles, we used the following criteria: we searched for the key words "Sensor", "Augmented Reality", "Training" and "Skill". We observed that the use of the additional keyword "expert/expertise" delivered fewer results. The electronic databases included in this review were: ScienceDirect (Elsevier), SAGE, ACM and SpringerLink. We also addressed only recent studies (since 2014–2016), because AR and sensors are developing technologies that have seen a recent spike both in investments and funding since 2015 (CBINSIGHTS, 2016) and in maturity for mobile AR support. Only studies that provided an explicit description of their prototypes, the sensors they used and the objective of the prototype were selected for evaluation. Fig. 2.1 illustrates a number of different steps at which the articles were screened.

We examined the abstracts of 268 studies, which we identified among the initial search results from the keyword search. We evaluated these studies to check whether they met our inclusion criteria. Out of 268 studies identified, only a total of 78 studies, which were composed of 26 studies from SpringerLink, 2 studies from SAGE, 21 studies from ACM and 36 studies from ScienceDirect, were considered relevant.

After we had selected relevant studies, we first examined each of them in detail to understand how the authors had captured the expert performance. We identified studies that captured expert performance either using sensors or performing task analysis or both. No distinction was made based on the type of involvement of the expert, to facilitate the inclusion of papers that may not have used recent sensors to automate the capturing of expert performance.

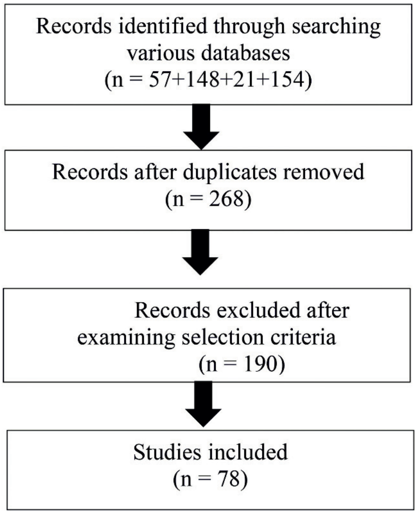


Figure 2.1: Process of finding the articles.

While doing so, we also observed and clustered common instructional methods employed by the studies. Instructional Design Methods (IDMs) are learning design patterns that leverage expert performance to support training using sensors and AR. IDMs are built on the concept of capturing the expert performance and using it to train apprentices. IDMs are abstract from the domain, and from other factors such as the particular vendor sensors. These IDMs have been evaluated to test their abstractness from the domain with the experts in the context of a European research project, Wearable Experience for Knowledge Intensive Training (WEKIT[1]).1 The WEKIT project explores the potential of sensor-based AR to train apprentices in three different domains, namely: maintenance, medicine and astronaut training. The experts used to evaluate the IDMs were experts in the domain mentioned and are experienced trainers. To validate the abstractness of IDMs, we selected three representative tasks, one from each domain, and conducted task analysis with the respective experts to identify the knowledge and skills required to perform these tasks. Later, the experts were provided with a list of IDMs along with their descriptions and asked to select the IDMs that could be used to train the skills identified in this task analysis. The results showed that most IDMs are applicable across all three domains, which validates the claim that IDMs are applicable independent of the domain.

To further analyse these studies, we categorised these IDMs based on the four components of the 4C/ID model. Based on the assumption that all types of learning task can be represented in the form of the four components: 1) learning task, 2) supportive information, 3) just in time information and 4) part-task practice (van Merriënboer et al., 2002), we aimed to explore whether the sensors and AR are capable of supporting the 4C/ID model-based training approach with the help of expert performance. The results of the analysis are presented in the following section.

[1]http://wekit.eu/

2.3 Results

2.3.1 Capture of expert performance

Sensors have the potential to capture various observable aspects of expert performance and the environment in which the expert performs. Some of these aspects may not be visible or obvious to the apprentice. Sensors can make invisible aspects of the task visible to apprentices, allowing them to achieve a better understanding of the process (Collins et al., 1991). In addition, recorded expert performance data can be used to train apprentices, but it can also assist experts by enabling them to quickly create training materials while demonstrating. Apprentices have the opportunity to replay the expert's demonstration in a much richer manner while actually doing the task or, when needed, to learn from the demonstration. However, expert performance data can also be used for other purposes, such as providing formative feedback (Schneider et al., 2017) by using expert performance data as a benchmark. Sensor systems can also read and log apprentices' performance, which can allow an expert to keep track of an apprentice's progress, enabling the expert to provide feedback when needed. To incorporate such functionalities into a system, it is crucial to understand how sensors can be used to capture expert performance. Therefore, we investigated how sensors have been used to capture expert performance in the studies collected. From our analysis of 78 studies, we identified 25 studies that exclusively modelled the expert (classified as Task Analysis in 2.1) or used sensors to record the expert's demonstration of the task (classified as Demonstration in Table 1). We have summarized them in Table 1, which consists of 4 columns: The approach column defines the type of data that was captured, while the second column defines the types of sensor used. This distinction was made to ensure that our results were not confined by the type and vendor of the sensors. There can be more than a single type of sensor to capture a particular set of data; for example, different sensors can be used to measure stress levels. The table also includes the IDMs the authors used to exploit the recorded expert performance. The Studies columns lists the studies from the literature that implemented the approaches. Finally, the last column lists the type of IDM such recording approaches support.

Table 2.1: Capture of expert performance.

Demonstration of the task

Approaches	**Sensor**	**Studies**	**Instructional Design Methods**
Recording of body movement	• Infrared depth camera	• (Wei et al., 2014) • (Khan, 2015) • (Prabhu et al., 2017)	• Augmented mirror • Formative feedback

Continued on next page

Table 2.1 – continued from previous page

Approaches	Sensor	Studies	Instructional Design Methods
Recording of hand movement	• Inertial camera • Infrared camera	• (Sun et al., 2017) • (Haug et al., 2014) • (Kowalewski et al., 2017) • (Ahmmad et al., 2014) • (Kritopoulou et al., 2016) • (Zhao et al., 2016)	• Haptic feedback • Augmented path • Formative feedback
Recording of force applied	• Pressure sensor	• (Araki et al., 2017) • (Asadipour et al., 2017)	• Contextual information • Haptic feedback
Recording of focus areas	• Eye tracker • Location tracker	• (Sanfilippo, 2017) • (Roads et al., 2016)	• Point of view video • Contextual information • Highlight object of interest • Directed focus
Modelling the task with task analysis			
Live telepresence: Expert provided audio instructions and visual guides during the live simulation	• Microphone • Video Camera	• (Sanfilippo, 2017) • (Li et al., 2015) • (Datcu et al., 2014) • (Chinthammit et al., 2014) • (Bordegoni et al., 2014)	• Contextual information • Point of view video • Annotations • Haptic feedback • Cues and clues • Directed focus • Audio Instructions
Modelling of the expert: Interviews with various experts		• (Schneider et al., 2017) • (Sebillo et al., 2015) • (Rozenblit et al., 2014) • (Meleiro et al., 2014) • (Djajadiningrat et al., 2016) • (Kim and Dey, 2016)	• Feedback • Haptic feedback • Contextual information • Haptic feedback • Haptic Interactive virtual objects
Assistance: Design of system to assist expert to train the apprentice		• (Daponte et al., 2014)	• Augmented Mirror

Table 2.2: List of instructional design methods.

IDMs	Description	Methods for capturing	Methods for enactment	Studies
Augmented Paths	Augmenting virtual path overlaid on the physical world in a way which allows the apprentice to guide his/her motion with precision	• Recording of body movement • Motion sensors • Depth camera	• Visualizing expert paths using AR display • Comparison to expert data • Feedback by capturing apprentice movement with inertial sensor	• (Liu et al., 2017) • (Stunt et al., 2016) • (Sun et al., 2017) • (Haug et al., 2014) • (Rozenblit et al., 2014) • (Tokuyasu et al., 2014) • (Kritopoulou et al., 2016)
Augmented Mirror	Augmented display where the apprentice can track his/her body, similar to dance rooms	• Recording of body movement • Infrared sensor	• Large display where the apprentice can see himself/herself • Infrared sensor to provide visual feedback	• (Chia and Saakes, 2014) • (Meleiro et al., 2014) • (Khan, 2015) • (Daponte et al., 2014) • (Lin et al., 2015) • (Wei et al., 2014)
Highlight Object of Interest	Highlights physical objects within the visual area, indicating to the trainee that the expert found that object of interest	• Recording of focus area • Video recording • Eye tracking sensor	• AR display to highlight objects • Eye tracker for formative feedback sensor	• (Roads et al., 2016) • (Tong et al., 2016) • (Sand et al., 2016) • (Rane et al., 2016)
Directed Focus	Visual pointer for expert-determined relevant objects outside the visual area	• Recording of focus area • Video recording • Eye tracking sensor • Location tracker	• AR display to show pointer • Eye tracker for formative feedback sensor • Location tracker	• (Ke et al., 2016) • (Zhu et al., 2014)
Point of View Video	Provides unique trainee/expert point of view video which may not be available in a third person perspective	• Head-mounted camera • Interaction mechanism to initiate and stop recording • Zoom in to subject	• AR display for unobtrusive viewing of video • Interaction and inference mechanism • Zoom in to video	• (Milazzo et al., 2016) • (Chinthammit et al., 2014) • (Bordegoni et al., 2014) • (Sanfilippo, 2017)
Cues & Clues	Cues and clues are pivots that trigger a solution search. They can be in the form of image or audio or any form of annotations. They should help the apprentice explore the solution with an annotation	• Record annotations • Take a picture, save video, audio or text • Use a physical object in the real world as an anchor for placing the cues and clues within the context • Location tracker	• Display on demand • Inference mechanism to automatically display the cues and clues • Location tracker	• (Datcu et al., 2014) • (Allain et al., 2015) • (Chang et al., 2015)
Annotations	Allow a physical object to be annotated by the expert during task demonstration (similar to sticky notes, but with more modes of information)	• Methods to tag media with a physical object • Can be done with audio, video or text • Location tracker	• AR display mechanism to read the annotations • Mechanism for unobtrusive relay of information • Location tracker	• (Datcu et al., 2014) • (Bordegoni et al., 2014) • (Li et al., 2015)
Object Enrichment	Provides domain-related information about the physical artefact which is crucial to the performance of the task from an expert's point of view		• Object recognition • Control over information displayed • Display information on demand	• (Wang et al., 2016) • (Zhu et al., 2014) • (See et al., 2016) • (Perlini et al., 2014)

Continued on next page

Table 2.1 – continued from previous page

IDMs	Description	Methods for capturing	Methods for enactment	Studies
Contextual Information	Provides information about the process that is frequently changing but is important for performance	• Record of the steps involved in the process • Recording of pivotal points that define the beginning and end of a task • Tagging of dynamic information such as location of the colleague in collaborative tasks, for tracking	• Method to know when and where to provide the information • Procedure information that depends on the context and is required for the task is presented when needed	• (Djajadiningrat et al., 2016) • (Wang et al., 2016) • (Sand et al., 2016) • (Rane et al., 2016) • (Zhu et al., 2014) • (Manuri et al., 2014) • (Kersten-Oertel et al., 2016) • (Kwon et al., 2014) • (Islam et al., 2016) • (Sano et al., 2016) • (Park et al., 2016) • (Dalle Mura et al., 2016) • (Sousa et al., 2016) • (Kim and Dey, 2016) • (Altimira et al., 2017) • (Such et al., 2014) • (Sebillo et al., 2015) • (Borges et al., 2016) • (Perlini et al., 2014)
3D Models and Animation	3D models and animations assist in the easy interpretation of complex models and phenomena which require high spatial processing ability	• Allow expert to place 3D models and animations in physical space and time	• AR display to display the model and animation • Interaction mechanism such as gestures to interact with the models where needed	• (De Ravé et al., 2016) • (Radu et al., 2015) • (Manuri et al., 2014) • (Lee et al., 2016) • (Cirulis and Liepina, 2014) • (Kamphuis et al., 2014)
Interactive Virtual Objects	Manipulatable virtual objects to interface with physical interactions for practice. Unlike 3D models and animation, which render a phenomenon, these objects' behaviour outcomes are determined by the specific physical interaction		• Ontological models where behaviours of the model are defined by the relationships and variables • Motion tracking with inertial sensors • Haptic and visual feedback	• (De Paolis et al., 2014) • (Liu et al., 2017) • (Jang et al., 2014) • (Ke et al., 2016) • (Radu et al., 2015) • (Juanes et al., 2015) • (Lok et al., 2014) • (Gallegos-Nieto et al., 2017) • (Lahanas et al., 2015) • (Lee et al., 2016) • (Benedetti et al., 2014) • (Onishi et al., 2014) • (Freschi et al., 2015) • (Oyekan et al., 2017) • (Kowalewski et al., 2017)

Continued on next page

Table 2.1 – continued from previous page

IDMs	Description	Methods for capturing	Methods for enactment	Studies
Haptic Feedback	Force feedback for perception and manipulation of authentic objects by means of a haptic sensor, to provide feedback and guidance	• Recording of hand movement • Recording of body movement • Inertial sensors • Infrared sensors	• Tracking the apprentice's motion with inertial sensor for comparison with expert performance • Motors for providing haptic feedback	• (Stunt et al., 2016) • (Haug et al., 2014) • (Gallegos-Nieto et al., 2017) • (Wang et al., 2016) • (Condino et al., 2016) • (Kim and Dey, 2016) • (Asadipour et al., 2017) • (Funk et al., 2016) • (Ahmmad et al., 2014) • (Araki et al., 2017)
X-ray Vision	Visualizing the internal process invisible to the eye for enhanced understanding. Overlaying internal workings of the physical system on the physical object itself provides enhanced understanding of the system		• Visualisation of the phenomena overlaid on the physical object • Interaction mechanisms • Object recognition	• (Kersten-Oertel et al., 2016) • (Kwon et al., 2014) • (Freschi et al., 2015) • (Buń et al., 2015) • (Condino et al., 2016) • (Cirulis and Liepina, 2014) • (Islam et al., 2016) • (Koreeda et al., 2016) • (Shekhar et al., 2010) • (Kamphuis et al., 2014) • (Chong et al., 2015) • (Choi et al., 2017)
Formative Feedback	Formative feedback is any feedback that can be provided by sensors and AR. It could be provided in visual, auditory or haptic form and should assist in conveying the procedural information	• Recording of any relevant sensor data in a meaningful manner such that it can be used for comparison with incoming data streams	• Measure of performance • Mechanism to evaluate the overall performance	• (Matassa and Morreale, 2016) • (Prabhu et al., 2017) • (Horeman et al., 2014)
Summative Feedback	Summative feedback is a versatile TM that is provided at the end of each practice session. It should allow reflection on the current performance	• Recording of any relevant sensor data in a meaningful manner such that it can be used for comparison of the final result of the expert and the apprentice	• Mechanism to assess the overall performance • Measure of overall performance	• (Zhao et al., 2016)
Mobile Control	Allows execution/visualization of remote action or controls which would otherwise require leaving the current workplace	• Task analysis to determine what actions and outputs are relevant • Implementation to control devices remotely	• Implementation to control physical objects manually	
Ghost Track	Allows visualisation of the earlier recording of the apprentice himself/herself for reflection	• Sensors to capture necessary data for reflection • Recording of the results of the action performed	• Visualisation mechanisms • Tracking of the current state of the apprentice	

2.3.2 Instructional Design Methods

We extracted the IDMs from the selected group of 78 studies. Our intention was to understand and document different IDMs that exploited the affordances of sensors and AR. We defined three general characteristics based on our observation of the implementation of IDMs, which are given in Table 2. Each IDM is characterised by a description, the unique features that define its implementation and the skill it can be used to train. The other characteristics include requirements for recording, such as hardware and software requirements, and for re-enacting by the apprentice, which may include wearable technologies. The list of questions is by no means exhaustive, as the IDMs are abstract from the domain, but using IDMs in the specific context typically leads to specialisations. In addition, requirements for recording are not necessarily bound to the recording approaches found in the studies selected (see 2.3).

From our analysis of the selected 78 studies, we identified 18 IDMs. The studies and the IDMs they implemented are mapped in Table 3, along with the means and methods for capturing and enacting expert performance. Each IDM is identified by its self-descriptive name in the first column from the left. After identifying all the IDMs, we further classified the IDMs according to the four components of the 4C/ID model in order to explore how sensor-based AR can be used to facilitate technology-enhanced 4C/ID learning environments. Evidence of the effectiveness of training environments designed in line with the specifications of the 4C/ID model for supporting deliberate practice in training contexts has also been documented by De Corte et al. (2003); Merrill (2002). In addition, Sarfo and Elen (2006) assessed technology-enhanced learning environments developed with 4C/ID specifications for expertise development and positively indicated that such systems have the potential to promote expertise development through deliberate practice.

Table 2.3: General Instructional Design Methods characteristics

<table>
<tr><th colspan="2">Description</th></tr>
<tr><td colspan="2">• How can the features be described?
• What skills are being addressed?</td></tr>
<tr><th>Requirements for Capture</th><th>Requirements for Enactment</th></tr>
<tr><td>• What types of sensors are required?
• What type of data must be captured?</td><td>• What sensor is required for enactment?
• What type of data is required enactment?
• How is this feature enabled by/for the apprentice?
• Which interaction means does the apprentice have?</td></tr>
</table>

2.3.3 Four component instructional design

The 4C/ID model is a non-linear and systematic processing model to design environments for complex learning. This model deals with complex skills by breaking down the complex task into sub-tasks without losing sight of the separate elements and the interconnections between them (Van Merriënboer et al., 2002). The assumption of the 4C/ID-model is that all complex skills and their sub skills can always be described in terms of four components, namely: 1) learning task 2) supportive information 3) just in time information, and 4) part-task practice. In the following section, we provide the results of the classification of the IDMs according to this framework. In addition, we have also mapped the type of attributes that the studies trained using the IDM.

Learning task

The IDMs in this component involve the apprentice in an active two-way interaction with the learning task itself. However, AR is usually used in the context of authentic learning, meaning that the learning task can be used either to support the authentic task or even to replace the authentic task in cases where the authentic task is not possible. Learning task IDMs are independent of supportive information and the procedural information. Table 4 lists the IDMs that support the learning task component. This component can be implemented with the IDMs in Table 4. For example, IDMs such as augmented path allow apprentices to imitate the expert performance for a smaller subset of the whole surgical complex skill.

Supportive information

IDMs in this component treat the apprentice as a non-dynamic receptor of supportive information that is not related to the procedure being executed. Supportive information is usually provided before the task execution, and during the task execution if needed. The "supportive information" section of Table 4 lists the IDMs that support this component by enabling the presentation of supportive information to the apprentice.

The supportive information component aims to elaborate the task model by establishing non-arbitrary relationships between the new elements and what apprentices already know. IDMs in this component, such as object enrichment, which enriches the physical object with virtual supportive information, allow apprentices to access domain-related information within the authentic context. This allows apprentices to link their existing mental model to its application during the training.

Just in time information

IDMs in this component treat the apprentice as a non-dynamic receptor of procedural information that is related to the procedure being executed. Table 4 lists the IDMs that have been identified in the "just in time" information component. AR has frequently been found to be well suited to providing procedural information in recurrent tasks, such as an assembly task. For example, an IDM such as haptic feedback is capable of providing corrective feedback precisely when needed.

Part task practice

Only IDMs that contribute to the repetition of the recurrent task are added in this component. However, we could only identify summative feedback as supporting this component. We have also listed Ghost track as an IDM that can support this component; this was not found in the literature, but was motivated by ghost records in games. This IDM can create ghost visualisations of a previous recording of an expert or the apprentice himself/herself, to provide a level at which the apprentice can aim in order to improve on the previous practice session.

To conclude, we have classified the IDMs that we identified based on the four components of 4C/ID, which provides us with an overview of how sensor-based AR can be used to create learning environments within the 4C/ID model. Such environments can improve the learning process between the expert and the apprentice, allowing the apprentice to learn better from the expert. Such systems also facilitate deliberate practice, which can potentially offer efficient support for the attainment of superior performance.

Table 2.4: IDMs classified based on the components of 4C/ID model and expert attributes.

IDMs	**Skills**			
	Motor	**Cognitive**	**Collaborative**	**Perceptual**
Learning task:				
Augmented Mirror	✓	✓		
Augmented Path	✓			
Interactive virtual objects	✓	✓	✓	
Mobile control		✓		
Supportive information:				
Object enrichment		✓		✓
3D models & animation		✓		
X-ray vision	✓	✓	✓	✓
Cues & clues		✓		✓
Annotation		✓		✓
Contextual information	✓	✓	✓	✓
POV videos	✓	✓	✓	✓
Procedural information:				
Directed focus		✓		
Highlight objects		✓		✓
Haptic feedback	✓	✓		✓
Formative feedback	✓	✓		✓
Part-task practice:				
Ghost Track	✓	✓		
Summative feedback	✓			

2.4 Discussion

Experts as mentors are crucial to the development of expertise in apprentices, but learning from experts has often proved to be a difficult task. Sensor-based AR introduces new possibilities for supporting apprentices and experts alike by capturing expert performance to train apprentices. Apprentices have the possibility to relive the expert's demonstration in a rich medium through additional sensor data augmented by AR. In addition, they can learn from the expert model at a more granular

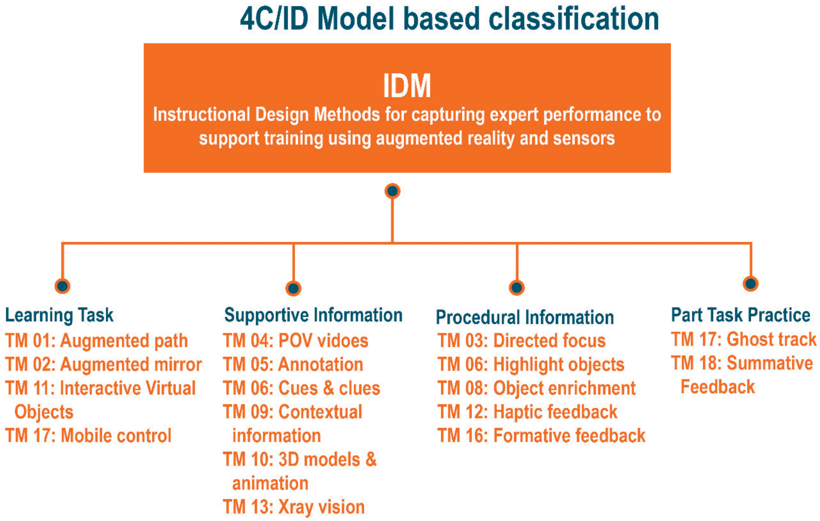


Figure 2.2: 4C/ID model-based classification of IDMs.

level, as experts may have overlooked details in their performance which could be crucial for the apprentice (Hinds, 1999). Capturing the expert performance and using it to train the apprentice helps address the shortage of experts to a certain degree, as an apprentice's performance can be compared with the expert model to provide formative feedback. Therefore, sensor-based AR posits potentials to simplify the process of learning from experts. To explore the use of sensors to capture expert performance and use sensor-based AR technology to train apprentices, we analysed 78 studies and identified 16 IDMs that use sensor-based AR for training apprentices based on expert data. The majority of IDMs are independent of the domain and applicable across domains. Furthermore, we classified the IDMs according to the components of the 4C/ID model to investigate whether sensor-based AR can support deliberate practice with the help of expert performance.

The exploration revealed that sensor-based AR has been used in various studies for training and supporting apprentices using experts as models. These methods varied from involving an expert in the system design and modelling an expert through various task analysis methods to using sensors to capture expert performance. In addition, we have also explored various types of expert performance attributes that can be captured with respective sensors. However, only 25 studies were identified that exclusively involved experts in their study. Seven approaches were identified from the 25 studies for capturing the expert performance. There is a need to emphasise more approaches to capture expert performance. Exploring new methods of capturing expert performance to allow apprentices to richly observe the expert demonstration could result in improved training efficiency. In addition, all the studies involved manual interpretation of recorded expert data between the capture and training phases. This increases the workload on the expert, which can result in hesitation to adopt the technology, as it is simply not feasible to edit the performance data manually for each learning task. To improve the learning process between the expert and the apprentice, a more automated approach needs to be developed which does not impose a significant additional workload on the expert.

It is worth mentioning that very few studies have exclusively claimed to address the cognitive process of the expert with the support of sensors. Cognitive aspects are not observable and cannot be directly captured by sensors. However, sensor data can be used to complement cognitive processes captured by performing a task analysis on the expert. For example, Jarodzka et al. (2013) used eye tracking data during the task analysis to explicate the cognitive process of the expert in order to understand the expert's action in classifying fish locomotion. Kim and Dey (2016) explored different physiological sensors such as eye tracking, EEG sensors and heart rate, concluding that there is a strong possibility that these sensors can also be used to represent the cognitive process. Further studies are required to explore how sensor-based AR can address cognitive aspects of training and expertise development. Our classification of IDMs (see Fig. 2) has revealed that the IDMs can be used for supporting deliberate practice within the 4C/ID model. We classified IDMs based on the four components of the 4C/ID model and on their intended pedagogic functionality. For example, haptic feedback was used to provide feedback on the procedures involved in performing the task (Wang et al., 2016), and thus was categorized under the procedural information component. Each component of the model consists of a list of IDMs which can be used to implement a sub-task in the 4C/ID model. In addition, we have also mapped the IDMs with the type of attribute each can be used to train. Performing task analysis on the complex task should yield one or more sub-tasks and the dominant attribute of the task. By selecting a proper set of IDMs that matches the attribute, a sensor-based AR training platform can be quickly designed.

While learning tasks in the 4C/ID model are arranged from lowest to highest difficulty to balance the task difficulty, the majority of studies do not address the affective aspects of tasks, which are crucial to deliberate practice. There are only a few exceptions, such as Altimira et al. (2017), who exclusively addressed affective aspects of the training by balancing difficulty in training to improve motivation. This could be an interesting area to look into for future studies; Bacca et al. (2014) have already emphasized the capability of AR to personalize training methods, reflecting the potential of sensor-based AR for affective aspects of the task. In addition, each study implemented and evaluated one or two IDMs together for their effectiveness in a concentrated application domain. When combining different IDMs, their effectiveness cannot be guaranteed, suggesting a need for further study of such a combination of IDMs. It should also be noted that most selected studies had a single domain of application, while we aim to improve the learning process between the expert and the apprentice regardless of their domain. Training tasks can often be very specific to the domain, while expertise is almost always specific to the domain of application. IDMs are abstract by definition, but designing a system following this approach may sometimes still require domain-specific implementations. To help facilitate this complex process, we have outlined a general operationalization guideline in Table 5 for designing 4C/ID-based learning environments.

Table 2.5: Steps for designing 4C/ID-based learning environments with IDMs.

No.	Steps	Description
0	Pre-design:	It is crucial to perform task analysis of the task to be taught by involving an expert of the domain. Task analysis can be done using interviews or other methods.
1	Design learning task:	Break the complex task into a set of sub-tasks and determine the performance attributes for each sub-task. Determine the required supportive information and procedural information for the particular task. If the sub-task is a non-routine task, it is best to adopt authentic scenarios, as they are better learnt in this manner. The selection of IDMs should be based on the dominant attribute of the sub-task.
2	Sequence the task:	As a complex task usually consists of more than one sub-task, it should be ordered in progression of increasing difficulty. It should be projected into the learning plan that when the apprentice finishes the last task in the list he/she will have mastered the task.
3	Determine performance objectives:	Criteria for allowing the trainee to progress to the next sub-task should be outlined. This also helps in focusing the type of feedback that can be provided.
4	Design supportive information:	Information that helps the apprentice perform the non-recurrent aspects of the sub-task are determined at this stage. This step may involve information that the expert will not be able to create during the recording of expert performance.
5	Record expert performance:	Based on the sub-task and its attributes, proper sets of IDMs from the learning task category are selected. Each IDM consists of a set of recording requirements which should be met. The expert proceeds after wearing all the sensors and begins to demonstrate the sub-task. The sensor records all the information and generates the expert performance data, which supports the procedural task. It should be noted that, when recurrent tasks are practised, procedural information should be scaffolded.
6	Train in same/similar environment:	It is crucial that re-enactment of the learning task is done in the same or a very similar environment. Technical requirements aside, it also helps the apprentice in his/her learning of the task without any overhead load.
7	Follow through reflection:	The system can provide formative and summative feedback based on expert data, but it cannot replace the expert. Providing the expert with logged data of the apprentice's performance in a simple, readable format can facilitate the learning process between the expert and the apprentice.

2.5 Conclusion

This paper presents the review of 78 studies that implemented sensor-based AR technology that were developed for training skills, found in literature between 2014 and 2016. Our literature study might have excluded studies using specific sensors, such as eye trackers, due to the keyword "sensor" not being adopted in the eye tracking community and the expertise study community having yet to fully embrace modern technology. Nevertheless, the analysis performed revealed sensor-based AR technology as a promising solution to support the training of apprentices using expert performance data, which also has the potential to support deliberate practice within the 4C/ID approach. Moreover, we are addressing the issues identified in learning from experts in vocational training scenarios. First, we supplement captured expert performance data for the apprentice to observe in a rich, multimodal manner whereby he/she may be able to understand and identify knowledge bits which the expert may have overlooked. Second, we provide the possibility to use the expert performance data for training by means of other mechanisms such as feedback. In this study, we focused on identifying state-of-the-art IDMs used by the prototypes in the selected studies. We did not analyse the IDMs identified according to their effectiveness or usability for training. Our main aim is to enhance the learning process between the expert and the apprentice. As such, we wanted to be independent of any domain-related results and implications. With young technology such as sensors and AR, which are only now beginning to mature, there is a need to explore ongoing research and the different creative outlets of the potentials that the technology can afford. Indeed, there is a need to evaluate the IDMs based on their effectiveness. However, rather than studying the effectiveness of the IDMs, we intended to see if the sensor-based AR can enhance the expert-apprentice learning process independently of the domain, and whether it can eventually support deliberate practice in apprentices.

The identified IDMs are promising and show high potential for effective implementation, as the sensor and AR technology will continue to improve in future. Research gaps such as the use of sensors to capture cognitive and affective aspects of expert performance provide an interesting area of study. While there are still many bottlenecks to be dealt with before the full potential of sensor-based AR technology can be realised, this literature review contributes an overview of the state of the art in instructional methods for using AR and sensor technology with expert performance. The study also touched on the potential of sensor-based AR for supporting deliberate practice. For our future research, we would like to use the identified IDMs to build new prototypes that embrace the 4C/ID model. The research group is currently developing generic capturing and AR learning components to be used in summative studies for the effectiveness, efficiency and usability of the IDMs.

Chapter 3

The ID4AR framework

With deliberate practice, the lumberjack's musical ability improved,
logarithmically !!

Chapter 3

The literature study revealed the lack proper methodologies for using augmented reality and sensors to facilitate deliberate practice of complex skills. However, it also established the potential of augmented reality and sensors to support training of complex skills with the four components instructional design (4C/ID) model. To address the lack of proper methodology for designing augmented reality environments for training complex skills, this chapter details the development of an abstract conceptual framework called Instructional design for augmented reality (ID4AR). ID4AR framework extends the 4C/ID model by encapsulating the 4C/ID model with AR based design patterns and hence, the affordances of augmented reality and senors, whilst providing the instructional designers with a methodological approach to design augmented reality training environments. It also provides guidelines and examples meant to help instructional designers use the ID4AR framework.

This chapter is published as: Limbu, B., Fominykh, M., Klemke, R., and Specht, M. (2019a). *A Conceptual Framework for Supporting Expertise Development with Augmented Reality and Wearable Sensors*, pages 213–228. Springer International Publishing, Cham

3.1 Introduction

Developing expertise is difficult for apprentices alone (Rikers et al., 2004). Ericsson et al. (2007) emphasise the importance of experts as mentors for supporting expertise development. However, experts tend to underestimate how difficult a task can be for apprentices (Hinds, 1999). Moreover, experts are often unaware of all the knowledge behind their superior performance (Patterson et al., 2010). Therefore, while experts are indispensable for expertise development in apprentices, learning from them is difficult. Limited access to the experts for apprentices also hinders their development even further. In order to mitigate these challenges, the Instructional Design for Augmented Reality (ID4AR) framework introduced in this paper aims to capture expert performance, making it accessible to many apprentices. By capturing expert performance as a resource, the ID4AR framework supports apprentices by emulating an expert-based guidance and feedback.

Sensors have the capability to unobtrusively measure physical properties. Wearable Sensors (WS) have been successfully used in training to provide feedback based on expert data (see, for example, (Jarodzka et al., 2013; Schneider et al., 2017)). A systematic review of literature and applications of WS and augmented reality (AR) posits a rich educational potential of these technologies for training (Bacca et al., 2014). A sensor and AR-based training environment with the expert recording can supplement training by providing guidance and feedback when expert is not available. The WEKIT project, in the context of which the work on the framework was performed, aims to exploit this potential of WS and AR based learning environment for supporting training using expert performance data.

AR provides a rich multimodal and multisensory medium (Azuma et al., 2001) for apprentices to observe the captured expert performance. Such a medium would enable apprentices to have access to expert data in authentic contexts when required. A key aspect of AR is to overlay the real world with virtual content to create an immersive platform (Bacca et al., 2014; Bower and Sturman, 2015) which places the apprentice in an authentic context while engaging all his/her senses. The affordances of AR and WS have the potential to supplement the expertise development in apprentices by using the captured expert performance (Guest et al., 2017). This has been reflected in the learning methodology adopted by the ID4AR framework.

The ID4AR framework is based on the learning methodology that aims to utilise the valuable experience and knowledge that the expert possesses with the help of AR and WS (Figure 3.1).

This learning methodology consists of three major phases: capturing expert performance, re-enacting expert performance by apprentices, and reflection (Fominykh et al., 2015). In addition, before the capturing phase, preparations are required to ensure that essential aspects of the expert performance are identified for capturing. The Capture phase ensures that the expert records all the relevant information needed for apprentices to perform the task. The re-enactment enables apprentices to learn from the recorded performance while the reflection phase allows the expert and the apprentice to reflect on the apprentice's performance by observation or/and from the data collected. The capture and re-enactment are supported by AR and WS. While AR and WS posit a rich educational potential for training, Bower and Sturman (2015) have emphasised putting pedagogy before technology. This is especially true for maturing technologies such as AR and WS which provide

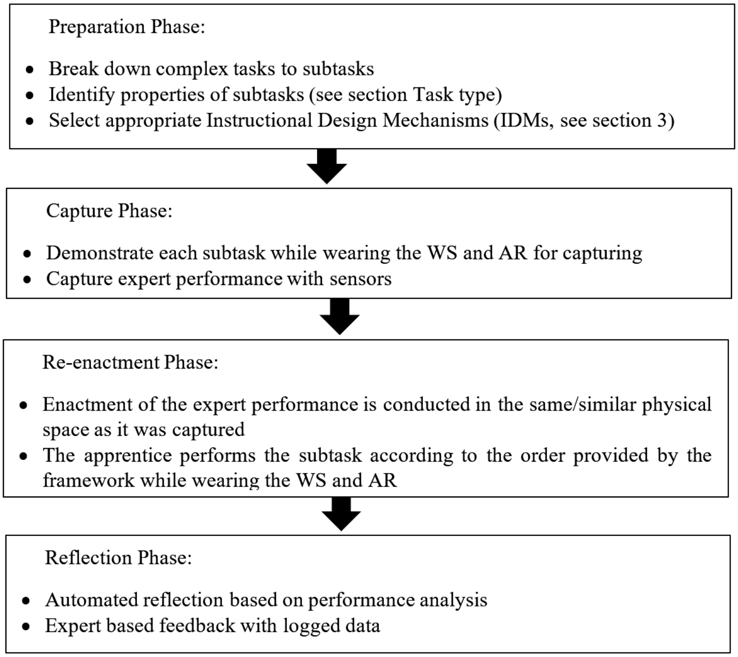


Figure 3.1: Phases of ID4AR framework learning methodology.

range of affordances potentially beneficial for training and education. Therefore, we structured the proposed framework around the pedagogical model known as Four Components Instructional Design (4C/ID) model.

Four component instructional design (4C/ID) model supports training of complex task for development of expertise (van Merriënboer and Kester, 2014). The 4C/ID model is a non-linear and systematic processing model for designing a complex learning environment. It is a holistic approach that decomposes the complex task into their simplest and smallest elements such that can be easily learnt by apprentices through a combination of these elements (Van Merriënboer et al., 2002).

3.2 Four Component Instructional Design

The 4C/ID (Four Component Instructional Design) model supports training of complex skills and has close resemblance with underlying principles of deliberate practice (Neelen and Kirschner, 2016). Deliberate practice is a focused practice on development of particular skill and is crucial for development of expertise. Sarfo and Elen (2006) assessed educational systems developed with 4C/ID specifications and positively indicated that the 4C/ID model promoted deliberate practice. Evidence about the effectiveness of training environments designed in line with specifications of the 4C/ID model for promoting deliberate practice in training contexts has also been documented by (Merriënboer and Paas, 2003) and Merrill (2002).

Sarfo and Elen (2006) reported positively that the 4C/ID model promoted development of expertise, which was based on their assessment of the technology enhanced learning environments developed with 4C/ID specifications. This claim is further backed by, in their study where they found that the 4C/ID model supports expertise development (Neelen and Kirschner, 2016). In addition, the sensors by personalising training in authentic contexts (Bacca et al., 2014) and the AR by supporting apprentices in real time, both facilitate deliberate practice and thus, eventually expertise development.

The ID4AR framework builds upon 4C/ID by facilitating the model with IDMs in order to support expertise development in apprentices with the help of AR and WS. By doing so, it bridges the pedagogic aspects of 4C/ID model with the affordances of AR and WS. The basic assumption of the 4C/ID model is that all complex learning can be represented in combination of four components (Learning Tasks, Supportive Information, Procedural Information and Part-Task Practice) described by the model (van Merriënboer and Kester, 2014). The ID4AR framework supplements the four components of the model with IDMs (Figure 2.2). In their turn, the IDMs support specific parts of training using AR and WS. Therefore, the ID4AR framework enables instructional designers to implement 4C/ID training using AR and WS. Figure 2.2 lists the IDMs that support each components of the 4C/ID model followed by brief description of its components.

3.2.1 Learning Task

Learning tasks are authentic, whole task experiences that are provided to the apprentice in order to promote schema construction for non-recurrent aspects of the task. For example, construction of schema by the apprentice can be facilitated by observation or imitation of the expert. The learning tasks, which are sub-task derived from the whole complex task, are administered in an increasing complexity and its dependency on other learning tasks. Each learning task is scaffolded to reduce the support and guidance when the apprentice attains higher form of expertise. In Figure 2.2, all the IDMs that actually allow apprentices to perform the task by imitating or observing the expert performance are placed under this component. It should be noted that this component overlaps often with part task practice, which emphasises on the repetition of learning tasks to enable automaticity. For the clarification sake, we will place the IDMs which support repetition aspect more in the part task practice component.

3.2.2 Supportive Information

Supportive information is the information provided to support schema construction, the learning and the performance of non-recurrent aspects of learning tasks, by supporting apprentices to deeply process the new information. The supportive information component aims to elaborate the whole task model by establishing non-arbitrary relationships between the new elements and what the apprentice already knows. Supportive information is usually provided before the task execution and during the task execution if needed which can be on demand or automated depending on the context. Figure 2.2 allocates all IDMs that provide domain level information for support as in supportive information component, compared to procedural information provided by the just in time component.

3.2.3 Procedural Information

Just in time procedural information is the prerequisite information to the learning and performance of recurrent aspects of learning tasks in a just in time fashion. AR has been frequently found to be well suited to provide procedural information in recurrent task such as an assembly task. In Figure 2.2, IDMs that assist in providing procedural information in a just in time fashion have been categorised under this section.

3.2.4 Part Task Practice

The last component of the 4C/ID model is the part task practice which recognises that some parts of the task are automatic and recurrent. Part-task practice items are provided to apprentices in order to promote rule automation for selected recurrent aspects of the complex task by means of "strengthening", in which cognitive rules accumulate higher strength on repeated successful executions (Van Merriënboer and Kirschner, 2017). All IDMs that facilitate repetition of learning task fall under this component.

The ID4AR framework was built upon the 4C/ID model as the foundation for the framework. The ID4AR framework supports the implementation of this learning methodology by collecting a pool of abstract AR and WS based instruction design methods or IDMs which can be applied to capture expert performance for training purposes. These IDMs are units of the framework that enables customisation of training platforms in various domains to meet the 4C/ID specifications.

3.3 Instructional Design Methods

The ID4AR framework is designed to be flexible enough to be used in different tasks. It manages to achieve this goal by building itself upon a pool of IDMs. IDMs are learning design methods that leverage on the expert performance to support expertise development using AR and WS. We do not claim to capture or explicate expertise, which is a complex notion in itself. By capturing relevant and measurable aspects of expert performance, we aim to support the development of expertise in the apprentices.

IDMs are abstract from the domain, and other factors such as the particular AR hardware and vendor sensors. The majority of IDMs were extracted from the literature by conducting a review of recent studies that exploited AR and WS for training (Limbu et al., 2018b). We identified three general characteristics of IDMs based on our observation of the implementation of IDMs which are outlined at Table 2.3. Each IDM is characterised by a description such as the type of skill it trains. The other characteristics include requirements for recording, such as hardware and software and requirements, and for re-enacting by the apprentice.

The list of IDMs identified is outlined in Table 2.2 along with their characteristics. IDMs require experts to demonstrate a task which allows sensors to capture his/her performance for creating expert model. The captured expert model can then be used for training. The list of IDM is not exhaustive and will only grow as technology improves. What the framework offers, is the insight on how to use these IDMs to capture expert performance for training. In the following section, we provide a guideline with example on how to operationalise the framework.

3.4 Operationalization of the framework

The ID4AR framework provides flexibility to adopt the ID4AR training approach to various training domains. The IDMs enables the trainer to select a proper set of IDMs for the current task being trained. The selection of the IDMs is based on the task attributes identified via extensive task analysis of the task to be performed. To facilitate the transition from task analysis to the application, IDMs have been categorised according to the skills which the authors of the original literature aimed to train using the IDM. The Table 2.4 provides classification of the IDMs with the attributes.

Table 2.4 assists the system designers and trainers to select the best set of IDMs for their use case. Once the IDMs are selected, the information in Table 2.3 can assist them to implement the system. However, before all this is done, the use case must be analysed to extract important attributes of the task. This can be done with the help of task analysis or a domain expert. The task can then be structured according to the frameworks 4C/ID approach for training. The list of steps, or guidelines are provided in the following section.

3.4.1 Guidelines/Steps to implementing the framework

The framework is designed to be abstract from the domain of application. Thus, it is crucial to perform task analysis of the task to be trained by involving an expert of the domain. Task analysis can be done using interviews or other methods. Below, we provide a set of guidelines to assist in implementing the framework.

1. Design learning task: Break the complex task into a set of sub task and determine the performance attributes such as mentioned in Figure 2.2, for each sub task. A sub-task is a fundamental task that constitutes the whole complex task and can represent a skill. Sub-tasks may be routine or non-routine. Routine task may benefit from IDMs in Learning task category such as interactive virtual objects however, authentic task should be preferred where possible. IT may be supplemented by IDMs in Part task section such as Ghost track for quick progress. Non-routine tasks are best left to authentic scenarios as they are better learnt in this manner.

2. Sequence the task: As complex task usually constitutes of more than one subtasks, it should be ordered in progression of increasing difficulty. However, the sequence of task should support variability of practice for better learning (Van Merriënboer et al., 2002). It should be projected into the learning plan that when the apprentice finishes the last task in the list he/she would have mastered the task.

3. Determine Performance Objectives: Criteria for allowing the apprentice to progress to the next sub-task should be outlined. This also helps in focusing the type of feedback that can be provided.

4. Design Supportive information: Information that help apprentices perform the non-recurrent aspects of the sub-task are determined. This step should generate contents that the expert will not be able to create or overlook during the recording of expert performance as non-recurrent task my not occur. Supportive information can be provided using one of the IDMs in the supportive information category, depending on the nature of information. Supportive information is usually only provided when requested so as not to over crowd the AR vision of the apprentice.

5. Record Expert performance: Based on the sub-task and its attributes, proper set of IDMs from the learning task category are selected. Each IDM consists of set of recording requirements which should be met. The expert proceeds after wearing all the sensors and beings to demonstrate the sub-task. The sensor records all the information and generates the learning content which supports the procedural task. It should be noted, when recurrent tasks are practised, procedural information should be scaffolded. Procedural information should be provided only when needed or requested during the practice.

6. Train in same/similar environment: It is crucial that re-enactment of the learning task is done in the same or closely similar environment. Technical requirements aside, it also helps in learning of the task by the apprentice without any overhead load.

7. Follow through reflection: The system can provide feedback on procedural task, but it does not replace the expert. Providing the expert with logged data of apprentices performance in a simple readable format will facilitate the learning process.

3.5 Operationalisation in WEKIT

This section is meant to provide an overview of how the framework is intended to be operationalised at the current state. We will present a use case scenario from the perspective of the framework as an example. The complex task of "Pre-flight inspection" task was broken into 10 sub-tasks after performing a through task analysis. In the first sub-task, "Ensuring that the baggage compartment is secured", task analysis performed revealed a set of attributes which will lead to proper selection of IDMs.

3.5.1 Task types

Perceptual ability is required in the first sub task of the pre-flight inspection task to be able to detect errors by means of observation. Similarly, High memory is also required to remember all the specifications regarding the task to be performed. In addition, in case of error detection, the technician is required to be able to cognitively analyse the situation. Experts also mentioned technicians are usually put through long hours resulting in fatigue. This may cause the technician to overlook details and thus they must be self-aware of their current state and their surroundings to avoid the risk associated with the task.

3.5.2 Instructional Design Methods

Based on the task types, "Directed focus" and "Point of View Video" was used to capture the perception of the trainer. "Contextual information" and "Think aloud protocol" was implemented to assist with memory. A checklist of the task needed to perform was provided for supporting the apprentice cognitively. Self-awareness was implemented with the help of biosensors and other sensors in the WEKIT prototype during the trial but is not accepted as IDM as it does not involve recording of the expert data. IDMs such as IDM "Formative Feedback" may be selected based on expert's opinion.

3.5.3 Capture

Each IDM possesses a set of recording requirements. After ensuring that all recording requirements are met, the expert will record the procedure ensuring that all the relevant information required for the re-enactment of the IDMs by the apprentice are recorded. Following a successful capture of data, the apprentice, with all the relevant information required to perform the task may initiate practice. Some information may not be available through the expert. Such information must be identified through the task analysis or through collective analysis of the sensor reading.

3.5.4 Re-enactment

The apprentice uses AR glasses and sensors which are used to project the captured data along time and space. Depending on the set of re-enactment requirements from the task types and IDMs, proper sensor set up is selected to track the apprentice performance. IDMs such as IDM "Formative Feedback" will provide lightweight feedback by using sensor readings.

3.5.5 Reflection

By comparing the expert performance with the apprentice performance, summative feedback may be provided. Comparison will be done between the current performance and earlier performance to facilitate self-reflection. The export will use the apprentice performance record to provide qualitative feedback.

3.6 Conclusion

The ID4AR framework attempts to provide a methodological approach to a newly emerging method of instruction using AR and sensors. With the technology rapidly developing, there is a need to formalise and explore methods and design for effective implementation of such technologies in learning context. AR and sensors are applicable in various domains and thus, with our framework approach, we defined an abstract methodology of designing training systems for vocational skill-based learning. The framework manages to utilise the full potential of the technology while being able to stay abstract from the tools used to perform the task. Similarly, the framework defines guidelines based on 4C/ID to ensure that the experts are being utilised to the full potential without compromising the training of the apprentice.

Eventually, the work done so far has presented potential and opportunities for further development and research. Even though several milestones have been met in the development of the framework, limitations exist. The system, with the current technological and research limitations will not be a substitute of the expert. The framework itself is designed to be a support for training where experts as resources are limited. The need to perform an extensive task analysis on the domain also exists. There is no evidence of explicating expertise and we do not claim to do so. While explicating the tacit knowledge is possible by rigorous manual means, by nature it cannot be done unobtrusively. Instead ID4AR will leverage on the performance metrics of the expert and visible attributes of expert performance to support the expertise development in the apprentice. The work on the framework is still ongoing. The list of IDMs is not exhaustive and will be updated as new findings and technology are revealed. IDMs and Task types will be more clearly defined to make the framework more concrete to meet 4C/ID specifications.

While the work is still in progress, many reflections have been made in the project life span. The proposed framework is effective for apprentices who are novice and need guidance at every step. Scaffolding may be applied in future practice sessions to help apprentices transition. However, apprentices at higher level of expertise learn differently requiring more cognitive aspects. The proposed framework relies on expertise demonstration for capturing data. Sensors are incapable of explicating cognitive expertise and these needs to be manually explicated. Similarly, many times the captured data needs to be manually tagged by the expert or algorithms specific to the domain is required to make use of the data. Therefore, we recommend using the model in earlier phases of learning to quickly attain certain level of expertise and transition into more self-monitored learning, if an expert is not available.

In conclusion, the ID4AR framework manages to facilitate training in skill-based learning, where apprenticeship is dominant. AR and WS systems designed with the framework will be able to address the shortage of experts and enable efficient attainment of expertise. The framework also assists in the design of the AR and WS based learning systems for technology enhanced learning with expert performance data.

Chapter 4

WEKIT.One: User study

So, what are your thoughts about the new AR glasses from Microsoft?

... I don't know, it kind of feels holo.

To address the lack of studies for facilitating deliberate practice of complex skills and to provide a methodological approach for designing augmented reality training environments, Instructional design for augmented reality (ID4AR) framework was built. The remaining part of this thesis focuses on validating the ID4AR framework by designing, developing and testing prototypes with it, following a design-based approach. This chapter reports on the design and development the WEKIT.One prototype along with the user study conducted in the three domains of the WEKIT project. The user study focuses on the validation of the implementation of instructional design patterns from the perspective of mentors and students.

This chapter is published as: Limbu, B. H., Jarodzka, H., Klemke, R., Wild, F., and Specht, M. (2018c). From AR to expertise: A user study of an augmented reality training to support expertise development. *Journal of Universal Computer Science*, 24(2):108–128. `http://www.jucs.org/jucs_24_2/from_ar_to_expertise`

4.1 Introduction

Sensors and augmented reality (AR) technologies have been developing fast with several plateaus of maturity being observed, such as with the release of Microsoft Hololens, over the past years. However, sensors and AR suffer from various constraints that obstruct their optimal implementation in industrial and educational contexts. Instructional design issues such as the distribution and flow of information between the physical and the virtual environment and between different devices is still obscure (Wu et al., 2013). As a consequence, designing a training environment based on sensors and AR for facilitating expertise development is challenging (Drljević et al., 2017). The complexity of interacting with large amounts of information and various devices at the same time, while performing a complex task can be overwhelming for the students. Designers of AR training environments need to realise the limitations to design the best possible training environment. In this regard, we adopt a design-based research approach (DiSessa and Cobb, 2004) which allows the end users to be a part of the design process ensuring that the final product meets the user requirements and needs. This article presents the first user study performed with our prototype designed for supporting expertise development in students in professional domains.

Attaining expertise is a difficult endeavour with claims that it may take up to 10 years to become an expert (Gladwell, 2008) . Ericsson et al. (2007) have emphasised the importance of experts as trainers for supporting expertise development. While trainers are imperative for supporting expertise development in an students, learning from them is difficult (Hinds, 1999; Patterson et al., 2010). Moreover, access to trainers is limited for students which impedes their development even further. Various efforts that can be potentially translated to address these problems have been made in the last years. Jarodzka et al. (2017) have presented eye tracking sensors as tools for supporting instructional design and expertise development in various domains such as chess and medicine. Similarly, posture sensors have been used for training public speaking skills by (Schneider et al., 2017). In addition, numerous similar studies have presented the potential of sensors and AR based training systems for expertise development (Olwal et al., 2008).

Sensors and AR have the capability to unobtrusively measure physical properties. The prototype used in this study utilises the potential of sensors and AR to record trainer's performance. This recorded performance is used to train students with the help of sensors and AR by making it available to the students when needed. By doing so with the help of Sensors and AR, the prototype supports technology enhanced training in authentic context which facilitates expertise development in students (Carey, 2014). In addition, sensors and AR also have the potential to provide personalised feedback and guidance in real time (Bacca et al., 2014). These potentials of sensors and AR are crucial aspects for supporting expertise development in students (Ericsson et al., 2007).

4.2 Sensors and AR for learning from trainers: the training methodology

To support expertise development, the training methodology of the prototype utilises the valuable experience and knowledge that the trainer possesses. It intends to make the experience and knowledge of the trainer accessible and available to students. This methodology consists of three major phases: recording trainer's performance, re-enacting trainer's performance by the student, and reflection. The recording phase ensures that the trainer records all the relevant information needed for a student to perform the task. The re-enactment enables the student to learn from the recorded performance while the reflection phase allows the trainer and the student to reflect on the performance by observation or/and from the data collected.

The training methodology is supposed to be applicable across different domains. To meet this criterion, we identified diverse attributes, such as speed and accuracy, important in various domains. Initially, a literature review and interview of the trainers in three professional domains, namely 1) aircraft maintenance 2) medical imaging and 3) astronaut training, were conducted to identify the attributes (Limbu et al., 2018b). IDMs used by the author in the studies reviewed, were extracted to support the training of each attribute (see Table 4.1). The commonly identified IDMs across all domains, as identified by the trainers in the domains, were implemented in the prototype used in this study. Table 4.1 provides the description of each of these IDMs along with how the trainer data was created which is outlined in the recording column. Similarly, the replay column describes how the trainer data was used for training.

Each IDM is mined from the literature review and the implementation is defined by the authors. IDMs defined in the context of this study, utilises recorded trainer's performance for training certain attributes of a skill with the help of sensors and AR. In contrast, the prototype described in this paper implements a pool of IDMs. Combining a pool of IDMs allows many inter-related aspects of a complex task to be trained. The prototype implements various IDMs together into a system which brings forth new challenges such as the usability of the system or even the assurance that each IDM implementation accomplishes its purpose. In addition, IDMs are abstract definitions and implementation methods can vary across platforms. Thus, we implement a collection of IDMs to test with end users in authentic settings and report the results of the first user study in this article.

Table 4.1: Implemented IDMs in the tested prototype.

IDMs	**Description**	**Methods of Recording**	**Methods of enactment**
Highlight Object of Interest	Highlight physical objects in the visual area indicating the students that the trainer found that object of interest	Interview with trainers to determine the object of interest	Hololens highlights the location of the object by using a virtual interface

Continued on next page

Table 4.1 – continued from previous page

IDMs	Description	Methods of Recording	Methods of enactment
Directed Focus	Visual pointer for trainer determined relevant objects outside the visual area	Interview with trainers to determine objects of interest and observation from demonstration	Visual direction indicated by an arrow to direct attention
Point of View Video	Provides unique student/trainer point of view video which may not be available in a third person perspective	Head mounted camera in Hololens are used to let the trainer record videos	Video projected by Hololens in the relevant physical location
Think aloud	Audio recordings the explanations and mental process (think aloud protocol) of the trainer during the task execution	Built in microphones in Hololens used to record trainer's explanations	Built in Hololens speaker plays the recorded audio in relevant time and space
Cues & Clues	Cues and clues are pivots that trigger solution search. It can be in form of image or audio. It should represent the solution with a single annotation	Place materials such as picture, audio etc. identified during interview with trainers to provide hints to the student	The chosen media contents are displayed using Hololens in the relevant time and location
Annotations	Allow a physical object to be annotated by the trainer during task execution. (Similar to sticky notes, but with more modes of information)	The trainer tags media to a physical object	Hololens displays the annotations by tracking the location of the physical objects
Object Enrichment	Provide domain related information about the physical artefact which are crucial to the performance of the task from a trainer's point of view	Interview with trainers determined relevant pieces of domain information apart from procedural information	Vuforia image recognition used to display such information in precise physical location
Contextual Information	Provide information about the process that is frequently changing but is important for performance.	Procedural information of the task is determined from the interview with trainer	Voice command based intractable checklist of steps to be performed is provided
3D Models and Animation	3d models and animations assist in easy interpretation of complex models and phenomena which require high spatial processing ability.	Modelling 3d object and creating 3d animation determined from interview with trainers.	Hololens display the 3d model which are move-able so that it is not obtrusive

Continued on next page

Table 4.1 – continued from previous page

IDMs	**Description**	**Methods of Recording**	**Methods of enactment**
Ghost track	Enables visualising the recorded movement of the person's whole body	Trainers body movement is recorded while demonstrating the task	Hololens enabled visualisation of recording through a holographic body enacting the recording.

In this study, we explore how the end users, that is the trainers and the students, from three professional domains perceive our prototype. To do so, we evaluate the prototypical implementation of the IDMs in three professional domains. It should be noted that the prototype is still in an early phase of development and thus measurements of effectiveness in training are not expected to be optimal. Therefore, in this study we test the following hypotheses:

1. System usability scale (Brooke, 2013) is at an acceptable level of 70 from both the trainers' and the students' perspective.

2. The prototypical implementation of each IDM will meet the authors' defined purpose of training certain attribute from the trainers' and the students' perspective.

3. The prototypical implementation of collection of IDMs will be equally accepted in all domains, both, from the trainers' and the students' perspective.

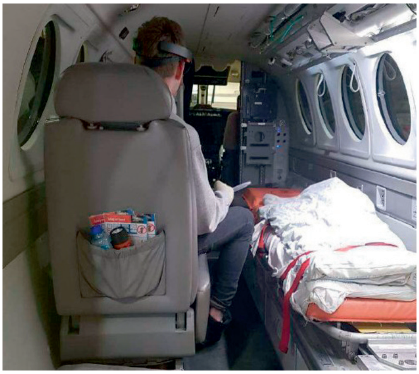
4.3 Method

4.3.1 Use cases and application domains

The prototype was tested in three different professional domains, namely 1) aircraft maintenance 2) medical imaging and 3) astronaut training. The aircraft maintenance training task consisted of ten steps of pre-flight inspection on an aircraft. Pre-flight inspection is used to determine if the aircraft is in an airworthy condition. Conducting a pre-flight inspection requires a lot of paperwork and reference information to be gathered and studied before proceeding to the aircraft to conduct the inspection. The ten steps in the pre-flight inspection require the participant to move along the aircraft cabin inspecting critical points for any hazards (see Figure 4.1a).

The medical imaging focused on the training of radiologist students to perform an echo-graphic examination by using an ultrasound machine. Unlike the aircraft maintenance, the participants are bound to a fixed location and the ultrasound machine which provides all the diagnostic information (see Figure 4.1b). The participants require operating knowledge of the machine, including the process of examining the patient with the ultrasound, and the perceptual abilities to recognise any deformities in the images produced by the machine.

Similarly, the astronaut training was conducted with the installation of the temporary stowage rack in the automated transfer vehicle. The procedure was performed on a mock up vehicle where the participants were required to install the rack using the proprietary installation units provided (see Figure 4.1c). The participants needed to know the location and the application procedure of the installation units which support the racks. Repetitive training sequences are needed to prepare the astronauts for all the activities and procedures required in space missions. These types of training practices accumulate to a large amount which takes significant proportion of training time.



(a) Participant performing pre-flight inspection inside the cabin

(b) Participant examining a patient using ultrasound machine.

(c) Installation of the rack by the participant for astronaut training domain

Figure 4.1: The application domains used in the study

4.3.2 Participants

The aircraft maintenance session in Lufttransport, Norway consisted of 31 students and 24 trainers. The students group comprised of student volunteers from bachelor programs of 'safety and environment', 'nautical sciences', and 'aviation' from the department of engineering & safety at The Arctic University of Norway. The trainers comprised of maintenance apprentices, skilled workers (mechanics) and technicians working in at Lufttransport. Similarly, 17 trainers and 22 students were involved in the astronaut training sessions in Altec, Italy. The trainers were Altec and Thales Alenia Space employees while the students were from the master in space exploration and development systems courses. During the medical imaging sessions in Ebit, Italy, 9 trainers varying from teachers to Medical doctors and 39 students from the faculty of medicine and ICT engineering participated in the session. Over all, in all three professional domains, there were 39 females and 103 males with most students age falling between 18-24 while most trainers age fell between 25-34.

4.3.3 Materials: AR and Sensor prototype

The prototype aims to utilise the valuable experience and knowledge that trainers possess and make it accessible and available to all students. To achieve this, the prototype consists of two major components: the recorder to capture trainer's performance and the player for supporting training of students (see Figure 4.2). The recorder ensures that the trainer records all the relevant information needed to support the training of students. The player enables students to learn from the captured performance.

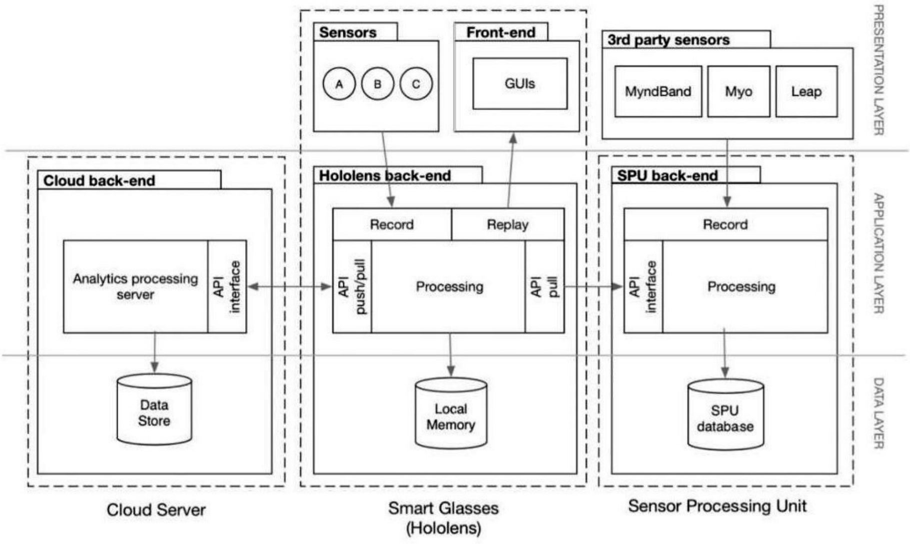
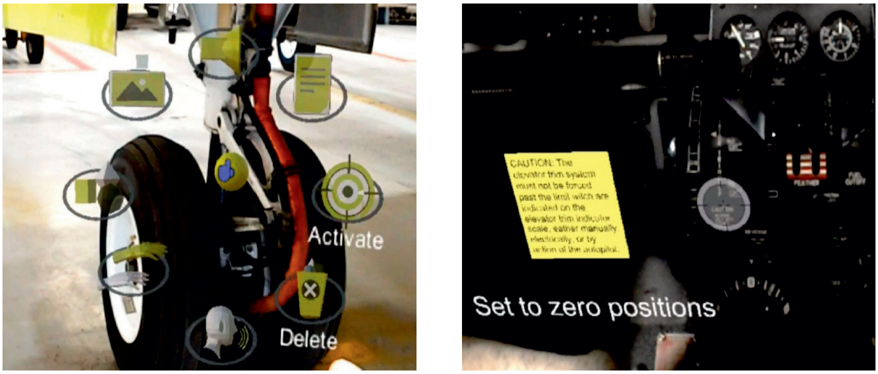


Figure 4.2: Architecture of the prototype

The prototype has been implemented for the Hololens which is an AR glass from Microsoft^{TM}, along with a combination of various wearable sensors. The recorder component uses various sensors depending on the requirement of the domain to record trainer's performance. It also allows trainers to create learning materials in authentic contexts as shown in Figure 4.3a. In Figure 6 the trainer can annotate the physical object with a virtual information. The trainer can interact with the recorder using gestures to annotate a physical location with various types of data. For example, the recorder allows trainers to record audio, take pictures and place

3D models at various physical locations. At the end of each recording, the data from the recorder is stored and fed into the player.

The player component on other the hand is catered for the students. The students receive step by step auditory and visual instructions which guides and supports them through the task. The contents created by the trainers such as notes are projected on their relevant physical locations and time based on the data from the recorder as shown in Figure 4.3b. students can also interact with the player using voice commands and gestures. The player allows students to navigate between the steps in the procedure using keyword based voice recognition. Both the recorder and the player are in the early stages of development. However, this study is more concerned with the usability of the system and the adherence of the implementation of IDMs according to the definition provided by the authors.



(a) Trainer's vision from the recorder for manually creating learning content

(b) Student's view of player captured from the Hololens.

Figure 4.3: Userview of the prototype

The prototype is developed in a three-layered architecture:

1. *Presentation layer*: the front-end and top-most level of the application, which consists of the graphical user interfaces (GUIs) and the sensor components to interact with the user and the external environment.

2. *Service layer*: the back-end and middle layer which coordinates the recorder and player clients, the data collection and analysis and the communication and transfer of these data across the platforms.

3. *Data layer*: the bottom layer where the information is stored such that it can be retrieved, processed, and re-presented to the user.

In addition to the three layers, the architecture combines three main computing units:

1. *Hololens*: The main wearable device through which the trainer can record his/her performance such that the learner can access it later. The Hololens will run both the two main applications of the prototype: the recorder and the player.

2. *Sensor Processing Unit (SPU)*: The portable computer device works as hub for the third-party sensors that are not embedded in the smart glasses but are necessary for capturing performance. The SPU is responsible only for the receiving and recording of all the third-party sensors. In addition, it also offers the necessary API interfaces to allow the Hololens to retrieve and store sensor data.

3. *Cloud Server*: The cloud-based server is the place in which the recorded performances are saved and processed for later re-enactment. The cloud-based solution allows for a scalable and distributed data storing over a nearly infinite number of computer nodes, as well as the availability of the data to all the connected and authorised devices.

4.3.4 Procedure

The participants were scheduled to arrive in a group of 2-4 participants per hour. They were initially introduced to the project and asked to sign a consent form. They were requested to fill in a demographic questionnaire prior to the evaluation study. Trainers were exposed to both the recorder and the player while the students were only exposed to the player. However, the students were informed of the scenario during the briefings that the content they saw was created by the trainers during the recording phase. Both the students and the trainers were familiarised with the user interactions on the Hololens by means of inbuilt gesture training in Hololens. After they completed the gesture training, the trainers were required to use the recorder under supervision to ensure that they were familiarised with the recorder. This session was followed by a briefing which involved only the trainers, on what they were expected to do. Finally, the trainers were asked to demonstrate the assigned tasks from their domain. A printed list of steps was also provided to the trainers for reference. After the recording, the trainers were briefed on the player aspects of the prototype. The students were not required to use the recorder. Instead, the students immediately exposed to the player after the gesture training. Finally, participants completed the questionnaire containing questions about the IDMs and system usability which was measured by Standard Usability Scale questionnaire (Brooke, 2013).

The IDM questionnaire evaluated the IDMs to measure their adherence to the intended definition of the IDM by the author. In the IDM questionnaire, students and trainers were asked to rate the statements on a Likert scale of 1-7 based on their experience after using the prototype. The participants rated these statements between completely agree and completely disagree based on their experience. The statements were derived from the description of each IDMs. Each statement represented an ideal experience of the implementation of the corresponding IDM. Similarly, to measure system usability SUS was used. SUS is an industry standard tool for measuring the system usability in a quick manner. The SUS scores calculated from individual questionnaires represent the system usability. SUS yields a single number between 0 to 100 (Brooke, 2013) representing a composite measure of the overall usability of the system being studied. Scores for individual items are not meaningful on their own. The acceptable SUS score is about 70 (Bangor et al., 2009; Brooke, 2013).

Sessions for each professional domain were held at their corresponding sites, with a week dedicated to each of them for preparation and execution. During the first session which is the aircraft maintenance, general technical issues and bugs in the prototype which affected the study directly were identified. These issues were resolved in the following sessions in case of astronaut training and medical imaging.

4.4 Results

4.4.1 System Usability Scale (SUS)

Table 4.2: SUS scores in all the sessions

	Aircraft maintenance	**Astronaut training**	**Medical imaging**
	M (SD)	M (SD)	M (SD)
Trainers	59.1(1.46)	69.2(1.06)	66.4(1.65)
Students	66.7(1.01)	67.5(1.83)	68(1.04)

The average SUS score for aircraft maintenance for trainers (59.1) is below 70 which indicates that the recorder's usability is not on an acceptable level yet. There is a noticeable improvement in the SUS scores of the recorder between trainers from aircraft maintenance session and the other two sessions (Table 2). Amendments made to the recorder after the first session i.e. aircraft maintenance may have resulted in the improved SUS scores for the recorder in astronaut training (69.2) and medical imaging (66.4) which are close to the acceptable value of 70. In addition, the operational difficulty of the prototype in the confined cabin space of the airplane caused usability issues such as difficulty to properly recognise gestures in dark places. The student's SUS score for the player in all sessions are close to the acceptable score of 70 (see Table 4.2).

4.4.2 Instructional Design Methods (IDMs)

Table 4.3: Average trainer ratings of the IDM items

IDMs	**Questionnaire Items**	**Aircraft maintenance**	**Astronaut training**	**Medical imaging**	**Average across all Domains**
		M (SD)	M (SD)	M (SD)	M (SD)
Directed Focus	*DF1.* I always knew where the next action happens	4.333 (0.730)	4.411 (1.175)	5.285 (1.112)	4.5 (0.707)
	DF2. I always knew where to stand and look	4.095 (1.374)	3.882 (1.053)	4.428 (0.786)	4.0 (0)
Highlight object of interest	*HL1.* I could always identify important objects	4.380 (0.864)	5.117 (0.992)	5.285 (0.487)	4.5 (0.707)

Continued on next page

Chapter 4

Table 4.3 – continued from previous page

IDMs	Questionnaire Items	Aircraft maintenance	Astronaut training	Medical imaging	Average across all Domains
Point of View Video	*POV1.* Videos provided a trainer's point of view on the task	4.666 (0.966)	5.352 (0.861)	5.833 (0.983)	5.0 (1.414)
Cues & Clues	*CUE1.* The floating photos helped me understand what the task	4.714 (0.956)	5.470 (0.799)	5.714 (0.951)	5.0 (1.414)
Annotations	*ANN1.* The virtual sticky notes helped me identify important bits of information	4.523 (0.928)	5.470 (0.799)	4.714 (1.112)	5.0 (1.414)
Object Enrichment	*OE1.* The system provided related information on objects of importance	4.809 (0.980)	5.764 (0.752)	5.248 (1.272)	4.5 (0.707)
3D models and animations	*ANI1.* The 3D animations helped me to interpret complex concepts	4.476 (0.872)	5.058 (0.747)	5.142 (1.214)	5.0 (1.414)
Think aloud	*TA2.* I understood what to do when following the trainer's audio recordings	4.761 (0.943)	5.470 (0.717)	5.714 (0.951)	5.0 (1.414)
	TA1. Audio recordings provided an trainer's explanations	4.761 (0.889)	5.529 (0.624)	5.714 (1.380)	5.0 (1.414)
Contextual Information	*CI1.* The system provided information relevant to the current situation and process	4.666 (0.912)	5.352 (1.114)	6.0 (0.577)	4.5 (0.707)
Ghost track	GT1. I was able to identify the position and the spatial orientation of the recorded trainer	4.619 (0.920)	5.176 (0.727)	5.428 (0.975)	5.0 (1.414)

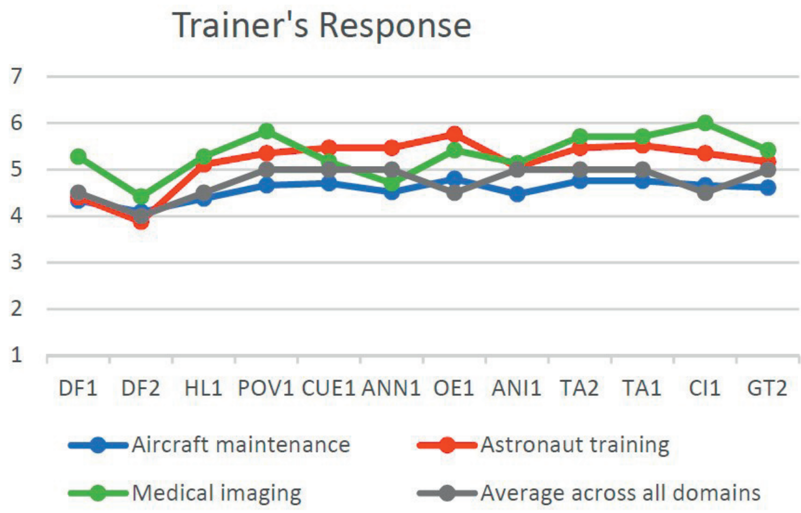


Figure 4.4: Trainer's response on the IDMs questionnaire

The trainer's ratings of IDMs across all three professional domains (see Section Use cases and application domains) is in general positive, ranking between 4 to 6, indicating positive acceptance of the implementation in all three professional domains (see Figure 4.4). Most IDMs such as Point of view (*M=5.0, SD=1.414*), Annotations (*M=5.0, SD=1.414*), Ghost Track (*M=5.0, SD=1.414*) etc. were rated above average by all the trainers in three sessions. Levene's test using the means (*p=0.169*) showed homogeneous variance between the three professional domains. In order to see if all the users of the three domains perceived the IDMs implementation equally which would hint that our prototype may be applicable across all 3 domains for training, we wanted to see if the differences between the results of the three domains were significant. Therefore, we conducted a MANOVA test and found a statistically significant difference in ratings between the three domains, $F\ (24,\ 62) = 1.587,\ p < .005$; $Wilk's\Lambda = .384$, $partial\eta^2 = .381$, mitigating the possibility that the results occurred by random occurrence. IDM Directed Focus was rated the lowest (*M=4.25*) across all the domains by the trainers. Details on the scores for each item can be found in Table 4.3.

Table 4.4: Averaged student ratings of the IDM items

IDMs	**Questionnaire Items**	**Aircraft maintenance**	**Astronaut training**	**Medical imaging**	**Average across all Domains**
		M (SD)	M (SD)	M (SD)	M (SD)
Directed Focus	*DF1.* I always knew where the next action happens	5.58 (1.104)	3.454 (1.710)	4.410 (1.292)	6.0 (0)

Continued on next page

Table 4.4 – continued from previous page

IDMs	Questionnaire Items	Aircraft maintenance	Astronaut training	Medical imaging	Average across all Domains
	DF2. I always knew where to stand and look	5.35 (1.368)	3.454 (1.595)	4.205 (1.293)	5.5 (0.707)
Highlight object of interest	*HL1.* I could always identify important objects	4.088 (1.147)	5.272 (1.695)	5.410 (1.044)	6.0 (0)
Point of View Video	*POV1.* Videos provided a trainer's point of view on the task	3.911 (1.815)	5.727 (1.279)	5.769 (0.916)	4.5 (2.121)
Cues & Clues	*CUE1.* The floating photos helped me understand what the task	2.617 (1.279)	5.772 (1.231)	5.666 (1.108)	3.5 (3.535)
Annotations	*ANN1.* The virtual sticky notes helped me identify important bits of information	4.970 (1.445)	5.545 (1.143)	5.564 (0.753)	6.5 (0.707)
Object Enrichment	*OE1.* The system provided related information on objects of importance	4.764 (1.327)	5.318 (1.170)	5.538 (0.853)	6.5 (0.707)
3D models and animations	*ANI1.* The 3D animations helped me to interpret complex concepts	5.058 (1.204)	5.272 (1.453)	5.589 (1.207)	6.0 (1.141)
Think aloud	*TA2.* I understood what to do when following the trainer's audio recordings	3.764 (1.102)	5.272 (1.241)	5.153 (1.159)	5.5 (0.707)
	TA1. Audio recordings provided an trainer's explanations	1.5 (1.022)	3.727 (1.723)	5.307 (1.217)	5.0 (1.414)
Contextual Information	*CI1.* The system provided information relevant to the current situation and process	3.941 (0.919)	5.818 (1.139)	5.512 (0.884)	5.0 (1.414)

Continued on next page

Table 4.4 – continued from previous page

IDMs	Questionnaire Items	Aircraft maintenance	Astronaut training	Medical imaging	Average across all Domains
Ghost track	GT1. I was able to identify the position and the spatial orientation of the recorded trainer	4.705 (1.030)	5.045 (1.252)	5.051 (1.050)	6.0 (0)

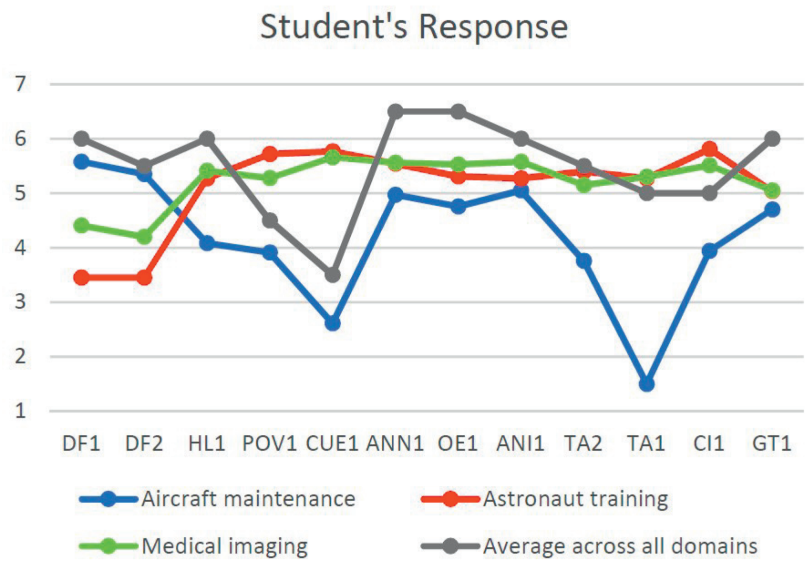


Figure 4.5: Student's response on the IDMs questionnaire

The overall students' ratings in all the three professional domains varied with scores ranging between 3 to 7 (see Figure 4.5). Levene's test using the means (*p=0.07*) showed a weak homogeneous variance between the three domains. This signifies that each domain perceives the IDMs differently. On conducting a MANOVA test, a statistically significant difference in ratings between the three domains, $F\ (24, 162) = 14.097$, $p < .005$; $Wilk's\Lambda = .105$, $partial\eta^2 = .68$ was found mitigating the possibility that the results occurred by random occurrence. The IDM Directed focus (*M=5.25*) was rated the highest in the aircraft maintenance domain while Contextual information (*M=5, SD=1.414*) was rated the highest by the students in Astronaut training. IDM Cue and clue was rated the lowest by students in the aircraft maintenance with score of (*M=2.617, SD=1.279*) despite being rated above average by the students in astronaut training (*M=5.772, SD=1.231*) and medical imaging (*M=5.666, SD=1.108*). The IDM Think a loud (*TA1: M=1.5, SD= 1.022*) and (*TA2: M=3.764, SD=0.102*) was rated the lowest by the students in aircraft maintenance. Nonetheless, on average across all three domains, most IDMs were generally accepted with ratings above average. Details on the score can be found in Table 4.4. In conclusion the results show that the usability of the prototype is close to meeting the first hypothesis. Similarly, the general acceptance of most of the IDMs by the trainers and the students rating the IDMs above the average value of 4, show that our implementation of the IDMs in the prototype meets the definition of the authors. In addition, since it was rated by students across three professional domains, our third hypothesis on interoperability across the domains is also met except for a few IDMs in some domains.

4.5 Discussion

The trainer rating of IDMs has been above average in all three domains with only slight difference from the first study conducted in the aircraft maintenance domain to the second in astronaut training and the third medical imaging (see Figure 4.4). The aircraft maintenance trainers rated the IDMs lower than the other 2 domains which may have been due to the study being carried out inside the cabin of the plane with limited lights and moving space which required participants to crouch all the time and Hololens to often lose tracking of the environment. Increasing the recorders dependency on physical markers with Vuforia may help to create a more stable AR experience. Regardless, the core implementation of IDMs was generally accepted by all trainers across the three domains. Therefore, future iterations of the prototype will only focus on implementing more IDMs and improving the overall experience for the trainer. The recorder's usability was improved based on the observations in the first session which accounted for the positive usability ratings in later sessions. Prior to implementation of the graphical icons based navigation, the recorder implemented a text based navigation. The trainers were required to aim the cursor my moving his/her head onto the text and then make a tapping gesture in order to select the menu. This was inconvenient in such a confined space as many missed taps were performed by the trainers. Hololens display are by nature opaque to a certain degree and reading smaller texts were difficult. Therefore, Graphical icons were implemented to make navigation easier and more intuitive for the trainers. Menu's that required text were made bolder and larger to make tapping easier. The learning curve for the trainers who are unfamiliar with technology was higher we well. In order to support such trainers, built in tool tip or help is required which will shorten the learning curve and allow them to quickly adapt the technology in their traditional training classes. Navigation indicators to show that the recording is being performed by the prototype, was implemented with icons turning red during the recording process. The recorder will also implement voice based navigation to help improve the usability of the system.

In addition, instruction sheets for the participants were also improved based on the experience from the first session which contributed to the overall experience. We also observed that the trainers mostly used audio recordings to create learning content due to its simplicity. It should be noted that the trainers were mainly exposed to the recorder. The recorder only records data required for the IDMs. Thus, the IDMs themselves are not implemented except some such as directed focus and object enrichment which are useful to the trainer as well. Though the trainers were also exposed to the player briefly, it was done so to allow them to get an understanding of how the data they recorded was being used. In addition, questionnaire related to the IDMs were, by nature, more oriented to the player and the students. The individual ratings may have also been affected by the short time frame each trainer was given. The learning curve may have been higher due to the complexity of operating the recorder on top of the complexity posed by a new technology such as AR.

The students' average ratings for all IDMs also increased in later sessions despite the core implementation of IDMs in the player not being changed between the sessions. However, the usability of the recorder was improved which led to the trainers recording better content for the students which may have improved the overall perception of students. The highest perceived ratings of the students in each of the sessions varied according to the domain. The significance of each IDM may have varied according to its perceived usefulness for the domain of use. For example, the directed focus was rated significantly lower by students from astronaut training and medical imaging while being rated higher in aircraft maintenance session. Directed focus may have been perceived higher due to larger work area in aircraft maintenance where the IDM provided significant advantage. In other sessions, the students did not have to move from a single fixed position to perform the task. At the same time, IDMs such as object enrichment and 3D models and animations were rated in an equal manner among all three sessions which could potentially hint that such IDMs can be applicable across all these three domains. In Figure 4.5, the think aloud protocol was rated with a significant difference between the two sub-questions. *TA2* asked if the participant understood reasoning behind the trainer's instructions as compared to *TA1* which only asked if they understood what to do next. Trainers in aircraft maintenance were limited by constraints such as time and physical space which may have affected their explanations. Experts or trainers in this case, tend to underestimate how difficult a task can be for the students (Hinds, 1999). Trainers are also often unaware of all the knowledge behind their superior performance (Patterson et al., 2010) and thus may omit the information an student would find valuable (Hinds, 1999). The largest pool of trainers in the aircraft maintenance session had limited time which did not allow each step to be comprehensively elaborated. Furthermore, it may have been due to the instructions not being explicit to the trainer, which was improved over the upcoming sessions. This is reflected in Figure 4.3a, where average ratings for the think-aloud protocol has improved in the latter sessions. IDM Cues and clues was rated the lowest across all three sessions, due to significant low ratings in the aircraft maintenance. It is unclear now as to why it was rated so and needs further analysis.

To summarise, this paper reports the first user study of the prototype designed to support expertise development utilising the recorded trainer performance data. The prototype is developed as a part of design based research project with forthcoming iterations in the future. Performing this study has provided us with a baseline for the measure of usability and a measure of proper implementation of IDMs. Combining various IDMs to enable support for different professional domains can generate many risk and challenges. Implementation of many IDMs may lead to increased complexity in the software and risk that each IDM implementation may fail to fulfil their purpose due to overhead in mixing various IDMs together. It is crucial to explore different approaches to design the system that reduces the learning curve, increases usability and overall achieves all the benefits of each implemented

IDM.

4.6 Conclusion

The implemented IDMs need to be better represented by the system before we can measure the learning outcome provided by the system. Based on observations, the recorder must implement other functionalities in a more intuitive manner reducing the learning curve for the trainers. This could otherwise limit the results of the future studies and the IDMs available to the students as the player depends on the recorded trainers' data. Both the students and trainers might also have been overwhelmed learning the new technology and range of functionalities implemented in the prototype in such a short time. To account for this, the system needs to be more intuitive. In addition, proper instructions can also be provided to the users along with more time allocated to each user to use the prototype. Due to the short time provided to the both the students and the trainers during the sessions, we were not able to collect much required qualitative data. Future studies may be focused more on smaller groups with more time for exposure. In addition, there are many difficulties trainers face to adapt the system in their regular training sessions as observed during the sessions. The system needs to support this transition to the best possible manner. It must also be complemented by proper instructions and training to support this transition.

Finally, more IDMs need to be implemented to support the domains more concretely. IDMs whose implementation were rated poorly will be further analysed and discussed with the trainers and the students to improve their implementation. The prototype used in this study was a linear system with minimal feedback being provided to the user. Proper feedback mechanisms will be implemented to enhance the usability and intuitiveness of the system. The usability of the system itself is not yet in an acceptable range. AR based usability guidelines will be further closely integrated to improve the usability in the system. Audio based interaction and proper user interface design to ease the learning of the system are some of the aspects that need improvement.

Chapter 5

WEKIT.One: Expert model evaluation

All scarecrows are experts because they are out standing in their fields.

Results from chapter 4 showed that the WEKIT.One application met the assumptions of the Instructional design for augmented reality (ID4AR) framework. With that established, this chapter explores the use of the WEKIT.one prototype for training. The WEKIT.One prototype relies on the mentor to create an expert model using it's recording functionality. This chapter reports on the evaluation of the expert model recorded with the WEKIT.one prototype to ensure that it meets the requirements for training.

This chapter is published as : Limbu, B., Vovk, A., Jarodzka, H., Klemke, R., Wild, F., and Specht, M. (2019b). Wekit.one: A sensor-based augmented reality system for experience capture and re-enactment. In Scheffel, M., Broisin, J., Pammer-Schindler, V., Ioannou, A., and Schneider, J., editors, *Transforming Learning with Meaningful Technologies*, pages 158–171, Cham. Springer International Publishing

5.1 Introduction

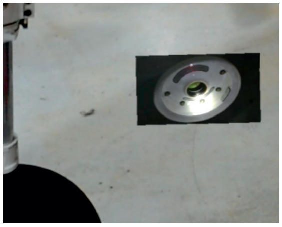
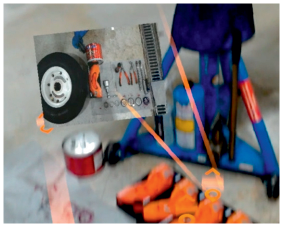
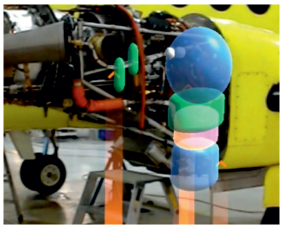
Augmented reality (AR) and sensors are becoming mainstream, also in professional technology enhanced learning and performance augmentation. Deploying AR for training, however, currently requires significant investment with regards to time and other resources, as most task-practice requires bespoke AR solutions. Arguably, the lack of standards and content to this day are one, if not *the* obstacle in the way of a widespread adoption, see (Langlotz et al., 2013), despite apparent benefits.

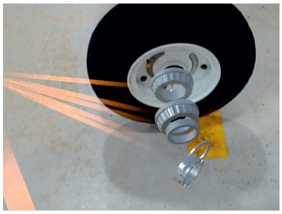
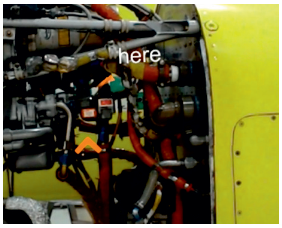
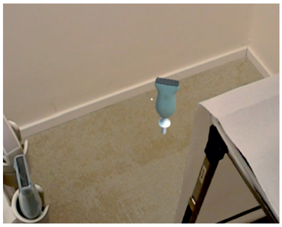
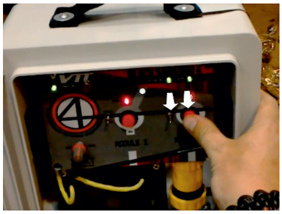
To mitigate this situation, we developed an abstract, domain-independent Instructional Design for AR (ID4AR) framework (Limbu et al., 2018b) in the WEKIT project, so as to foster adoption across different training domains. The model is designed to help reduce associated entry costs by providing the theoretical foundation and practical instructional design building blocks, so-called instructional design methods (IDMs), required to design and deploy AR and senor-based training applications. ID4AR includes a systematic collection of domain independent instructional design methods (IDMs) as its unit component. IDMs are based on the study of affordances of AR and wearable sensors and are also independent of hardware (and sensor) choice. Each IDM relies on recorded expert performance and performance-relevant data in order to support training with AR and a wearable sensors. The framework, which is rooted in the 4CID model for learning complex tasks (Limbu et al., 2018b,a), also supports instructional designers in the selection of required IDMs to meet the requirements of the intended solution. In addition, the framework defines, systematically, all procedures needed to record and replay such expert data. By satisfying the framework‘s requirements, instructional designers can more easily design complex AR and wearable sensor solutions for training. This paper provides the validation of this theoretical framework as domain independent tool for supporting instructional designers. To do so, WEKIT solution was developed using ID4AR framework which was used in all three professional domains of aircraft maintenance, medical imaging,and astronaut training.

The WEKIT solution (also called WEKIT.One) supports recording experts performance for efficient and in-situ authoring of learning materials. The solution, to cater to all three domains mention above , implements common IDMs found across all three domains, which were selected after extensive task analysis with the experts from the three domains. This was done to meet the time and resource constraints, instead of creating three different applications for each domain. While, (Limbu et al., 2018c) used domain experts to review the solution's compliance to the framework, this study investigates whether the solution can in fact be used to record expert models across the different domains. By deploying the solution in three different domains and evaluating the expert model created, we can draw back conclusions on the validity of the framework and its utility to design AR and sensor based solutions regardless of their application domain. Thus, in this paper we aim to examine: Are recorded expert performances from ID4AR based solutions fit to be used as expert models for training in all three domains?

To do so, we asked expert peers to evaluate the expert model according to their fitness for training. In addition, we also conducted a knowledge assessment study with students to validate that the model captured with the solution does not impact negatively on their learning, or, ideally, even improves in areas. In this paper, we present results of this expert-peer evaluation and the students knowledge assessment study which assessed the expert model recorded by the WEKIT solution built with the ID4AR framework.

Table 5.1: List of IDMs in WEKIT application.

IDM	Description	Visuals
Directed focus	Visual pointer for relevant objects outside the visual area of the trainee.	
Point of view video	Provides expert point-of-view video which may provide perspectives not available in a third person.	
Annotations	Allow a physical object to be annotated by the expert during task execution (similar to sticky notes but with more modalities).	
Ghost track	Allows visualisation of the whole-body movement of the expert or the earlier recording of the trainees themselves for imitation and reflection.	
Highlight objects of interest	Highlight physical objects in the visual area indicating to the trainee that the expert marked it as an object of interest.	

IDM	Description	Visuals
Object enrichment	Virtually amplify the effect of the process to enable trainees to understand the consequences of certain events or actions in the process which may be too subtle to notice.	
Contextual information	Provide information about the process that is frequently changing but is important for performance.	
3D models and animation	3d models and animations assist in easy interpretation of Complex models and phenomena which require high spatial processing ability.	
Interactive virtual objects	Interactable virtual objects to practice with physical interactions relying on the 3d models and animation.	
Cues and clues	Cues and clues are pivots that trigger solution search. They can be in any form of media but should represent the solution search with a single annotation.	
Haptic feedback	Lightweight force feedback for perception and manipulation of authentic objects by means of haptic sensor, to provide feedback and guidance.	

5.2 Method

To capture and evaluate the expert models, 61 experts and 337 students used the WEKIT solution during WEKIT trials held at Lufttransport in Norway for the aircraft maintenance, Ebit in Italy for medical imaging and Altec in Italy for astronaut

Table 5.2: Demographics for individual domains.

Domain		Gender			Experience			Trainers
	N	M	F	Age Range	<5	5 - 10	>10	N
Astronaut	13	11	2	25-34	2	4	7	2
Medical	26	18	8	25-54	13	0	13	2
Aeronautics	22	18	4	35-44	5	5	12	4
Total	61	47	14		20	9	32	8

training. These trails were conducted in a time span of more three months independently by the above mentioned use case organisations with out any intervention by other researchers and technical partners.

5.2.1 Participants

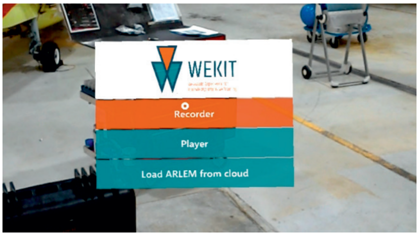
61 experts participated in the study from three different domains. The expert participants were defined as those who had experience in the domain they took part in. There were 47 male and 14 female expert participants, with the majority of them falling in the age range of 25-44. Among these participants, there were 8 supervisor, 8 trainers, 31 engineers and 19 from several other roles. 32 expert participants had more than 10 years of experience, 20 had less than 5 years and 9 between 5-10 years. Demographics for individual domains are detailed in Table 5.2

5.2.2 Apparatus

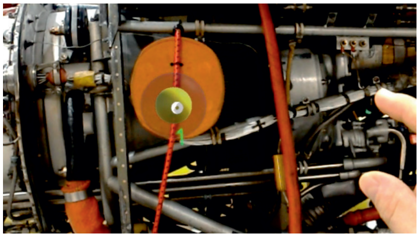
The WEKIT solution is built for the Microsoft Hololens, an AR platform. It is developed with Unity3D, Vuforia (marker-based image recognition toolkit for AR), and the Microsoft Mixed Reality Toolkit. The application consists of two main interfaces: the Recorder interface and the Player interface (see Figure 5.1a).

Recorder Interface

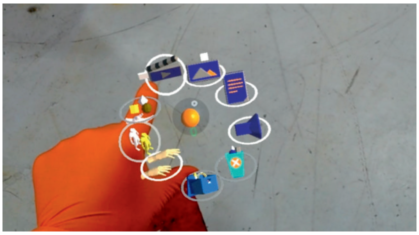
The recorder interface supports experts in creating learning content with two main functionalities: annotation of objects and locations in the physical space (using text, image, video, audio, 3D object annotations) and more implicit, observation-based, multi-modal capture of the expert performance, using sensor data. It provides two different methods of connecting virtual annotations to the physical space: marker-based and anchor-based. The marker-based approach relies on prepared image targets (using Vuforia for tracking), which binds augmented content to the physical environment to place the attached annotations relative to the marker image. The anchor-based approach uses the infrared scanner of the smart glasses to generate a spatial map of the environment to then attach all augmented content relative to physical anchor-points. Experts create so-called 'task stations' to record the learning activity in a systematic manner. Task stations can be placed by pointing the gaze cursor to the desired location and then performing a double-tap gesture, or by sticking the pre-trained image target marker onto an object or location (see Figure 5.1b). Task stations and their attached annotations are then subsequently translated to a linear or branched sequence of action steps in the player interface. Recorded units typically contain a longer sequence of such task stations (see Figure 5.2a), each typically with a combination of annotations attached. Experts can enrich the physical space with virtual images, point-of-view videos, voice recordings, place 3D models, mark the physical location as a point of interest, and record sensor data (see Figure 5.1c). The annotations working with sensor data currently make use of hand position, relative orientation (relative to the device), and the head position and orientation (relative to physical environment). Captured learning activities can be saved in the ARLEM format and can be uploaded to a cloud repository, when complete. ARLEM standard specifies how to represent activities for training knowledge, skills, and other abilities in a standardised interchange format for AR applications.



(a) Recorder and Player interface.



(b) Creating task stations.



(c) List of annotations.

(d) Sequence of task stations.

Figure 5.1: Recorder Interface

Player Interface

The player interface allows trainees to learn from the experts created learning contents. Students can download a learning activity from the cloud. Once downloaded, the player interface generates the user interface as a task list and task cards for stepwise guidance (see Figure 5.2a. The player interface projects the augmentations at the right location and in the right sequence (see Figure 5.2b). Students can navigate between the steps using voice commands or gestures.

5.2.3 Materials and measures

We aimed to evaluate the expert model's validity based on the recorded performances. Experts were considered to be experienced or working in the domain of the test-bed. For the actual evaluation, first, an expert performance was recorded in all three domains, producing three different models. These models were not post-processed. Second, the model was loaded and used by the peer experts according to

(a) Steps in a recording. (b) Content of a step.

Figure 5.2: Player Interface

their respective domains. Third, the peer experts evaluated the model using a specific questionnaire, i.e., the expert model evaluation questionnaire (EMEQ) which was the same for all three domains. Aim of the questionnaire, which is based on (Jucks et al., 2007), was to assess the characteristics of the expert model by judging its fitness for training. Participants responded by scoring questionnaire items on a Likert scale from 1 (=strongly disagree) to 7 (=strongly agree). The responses were collected through LimeSurvey, an online survey tool.

5.2.4 Design and procedure

The expert who captured the model was introduced to the WEKIT solution's user manual first to ensure he/she was familiar with the solution (e.g., using a generic gesture training). Prior to the recording, the expert was asked to plan the action steps and accordingly, the task stations with the affordances of the solution in mind. These included considerations such as how many task stations need to be created and what type of content would be presented in each of the task stations. During the subsequent recording of the activity, the expert was free to ask support questions. The expert was allowed to repeat the capturing process until satisfied. The peers who evaluated the model used both recorder and player. They used the recorder to understand how the model was created. In the player, the model that was initially created was loaded, and the peers followed through all the steps. The peers were also given as much time as they requested for the whole procedure. In the end, all the expert peers filled the questionnaire for evaluating the expert model.

5.3 Results

At the end of the three months duration of the WEKIT trials, the data was downloaded from LimeSurvey. In the following, we present the overall results and the results per domain. The mean response for the items across all three domains is presented in Table 5.3.

Table 5.3: Descriptive statistics for all three domains.

Descriptive statistics for EMEQ				
Items	Description	N	M	SD
EMEQ 1	It is important that the student knows what each key concept means.	61	6.066	.834

Continuation of Table 5.3				
Items	Description	N	M	SD
EMEQ 2	For this student, all key concepts are defined just in time.	61	5.574	.884
EMEQ 3	For this student, the procedure is explained in comprehensible enough terms.	61	5.198	.781
EMEQ 4	For this student, the procedure is explained in enough detail.	61	5.705	.882
EMEQ 5	All the information that the student needs to follow the procedure is contained.	61	5.852	.813
EMEQ 6	All the information that the student needs to follow the procedure is provided just in time.	61	5.574	.991
EMEQ 7	All the contained information is important to the student.	61	5.787	.951
EMEQ 8	All the information provided is non-obtrusive for the student.	61	5.639	.967
EMEQ 9	All the objects/items required by the student in the procedure is easily located/identified	61	5.577	1.203
EMEQ 10	It is clear for the student which physical area to move next.	61	5.459	.993
EMEQ 11	All relevant information that is frequently updated, such as temperature, is made aware to the student.	61	4.787	1.171

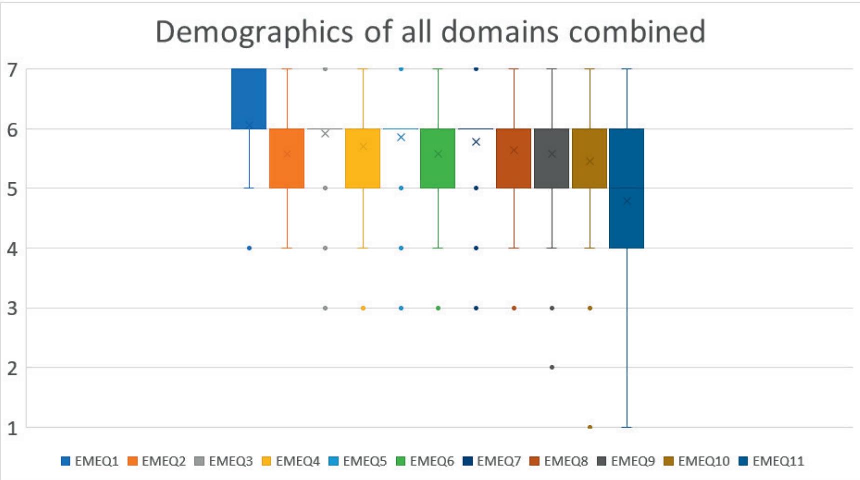


Figure 5.3: Demographics of all domain together

The average mean and the median response of experts across all trials for all the items were above average (see Figure 5.3). Experts strongly agreed on EMEQ 1 (Mdn = 6.07), on the importance for the students to understand what each key concept meant. Similarly, there is an agreement between expert participants for EMEQ 2 (Mdn = 5.57), EMEQ 4 (Mdn = 5.70) and EMEQ 6 (Mdn = 5.57) which verifies that the expert model explained the procedure in comprehensible terms and included all important information required for the procedure. Most expert participants had high degree of agreement in EMEQ 3 (Mdn = 5.92), EMEQ 5 (Mdn = 5.85) and EMEQ 7 (Mdn = 5.79). The procedure was found to have been explained in enough details, just in time and in an unobtrusive manner by the expert participants. The expert participants also found that the model guided students to the correct location and items in the physical space which was shown by EMEQ 9 (Mdn = 5.57) & EMEQ 10 (Mdn = 5.46). The SD of EMEQ 9 and EMEQ 11 was higher than acceptable. EMEQ 11 was rated between 1-7 with lower quartile rating the item between 1-4. Experts opinion vary hugely in terms of how well and often critical dynamic information were updated. Results of the study for individual domains are presented below.

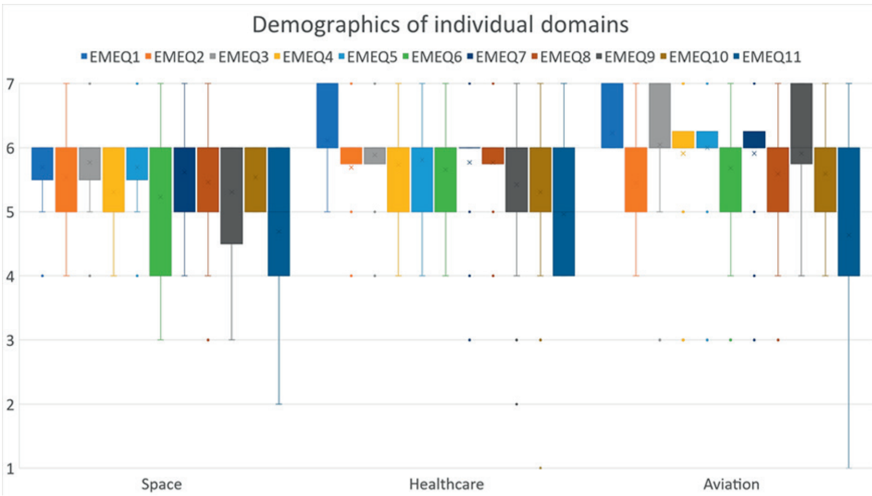


Figure 5.4: Demographics of study in individual domains

5.3.1 Astronaut domain

The expert participants mostly responded positively to the model with the median of all items above 4 (see Figure 5.3). Most expert participants responded positively on EMEQ 1 (Mdn=5.692) with only 1 participant rating it 4. For EMEQ 2 (Mdn=5.538), with the upper quartile between 6-7 and lower quartile between 4-5. This supported that most concepts were defined just in time in the expert model. Expert participants had a high level of agreement on EMEQ 3 (Mdn = 5.769) with only 1 expert participant rating it 4. Item EMEQ 4 (Mdn = 5.307), show that the expert model explained the procedure in comprehensible terms and details. Results of item EMEQ 5 (Mdn = 5.692) show that the contained information in the expert model is complete. EMEQ 6 (Mdn = 5.230) showed larger variation in expert participants agreement in terms of if the information was provided in the right time. EMEQ 7 (Mdn = 5.615) and EMEQ 8 (Mdn = 5.461) validates that the expert model contains all the important information, which are presented in an unobtrusive manner. Only 1 participant rating EMEQ 8 below 4. Item EMEQ 9 (Mdn = 5.307) verified that the participants were fairly able to locate the objects required for the procedure most of the time. The participants were also able to identify the place where the next step of the procedure was to be done. This is shown by the strong agreement between the expert participants in item EMEQ 10 (Mdn = 5.538). EMEQ 11 (Mdn = 4.692) showed loose agreement between expert participants with 75% rating it between 4-6 and the rest 25% voting it between 2-4.

5.3.2 Medical domain

The median response of experts for each item is above 4 (see Figure 5.3). There is consistent agreement among the expert participants for EMEQ 1 (Mdn=6.12), emphasizing that the students need to know the key concepts. While the majority of the expert participants think that most terms have been defined and in comprehensive manner, EMEQ 2 (Mdn = 5.69) & EMEQ 3 (Mdn = 5.75), 5 expert participants rated EMEQ 2 and 1 participant rated EMEQ 3 as 4. EMEQ 4 (Mdn = 5.73), EMEQ 5 (Mdn = 5.81) and EMEQ 6 (5.65) validate that the expert model was contained complete information which was provided just in time for the students. For all these three tiems, the middle quartile fell between 5-6. With only one expert participant rating EMEQ 7 (Mdn=5.77) and EMEQ 8 (Mdn =5.75) below 4, it can be argued that all information contained in the expert model were important for the procedure and were not presented obtrusively. The students were able to find the objects in the work space and were able to pinpoint the location for the next step in the procedure as shown by EMEQ 9 (Mdn = 5.42) and EMEQ 10 (Mdn = 5.31), with only 2 expert participants each rating them below 4. EMEQ 11 (Mdn = 4.96) shows that the relevant types of information were updated. However, 11 of the expert peers rated it at 4.

5.3.3 Aeronautics domain

The median response of the expert participants in this domain for each item are above 4. As with other domains expert in Aeronautics domain experts strongly agree that students should know what each key concept means, which is shown by EMEQ 1 (Mdn = 6.23). EMEQ 2 (Mdn = 5.45) shows that most key concepts were well defined in the model. 75% participants rated the item EMEQ 3 (Mdn = 6.05) between 6-7, with only 1 participant rating it 3. This shows that the experts found the model was comprehensible enough. Only 2 expert participants rated EMEQ 4 (Mdn = 5.91) below 4, which validated that the expert model was explained in enough detail. Similarly only one expert participant rated EMEQ 5 (Mdn = 6.00) below 4, with a strong agreement among the other expert participants which showed that the expert model contained all the information that the student needed to follow the procedure. EMEQ 6 (Mdn = 5.68) shows that the expert participants found that the information needed were provided in just in time fashion. All contained information was found to be important to the student in EMEQ 7 (Mdn = 5.91) , with only one expert disagreeing with a score of 3. EMEQ 8 (Mdn = 5.59) validates that the expert model was fairly unobtrusive for the students. The recorded model was also able to direct the participants to the location of the object required during the procedure most of the time as shown by EMEQ 9 (Mdn = 5.75). Similarly, EMEQ 10 (Mdn = 5.59) showed agreement among the participants that the students were provided guidance to move from one place to another during the procedure. Expert participants were divided for EMEQ 11 (Mdn= 4.64) which was rated 4 by 11 people, with distribution varying wildly from 1-7. The central quartile falls between 4-6.

One-way ANCOVA was conducted to determine any statistically significant difference between the three test-beds on EMEQ items. There is no significant effect of the application domain on EMEQ 1 [F(1, 59)=3.126, P=.082], EMEQ 2 [F(1, 59)=.175, P=.667], EMEQ 3 [F(1, 59)=1.905, P=.300], EMEQ 4 [F(1, 59)=3.720, P=.059], EMEQ 5 [F(1, 59)=1.281, P=.262], EMEQ 6 [F(1, 59)=1.423, P=.238], EMEQ 7 [F(1, 59)=.792, P=.377], EMEQ 8 [F(1, 59)=.049, P=.826], EMEQ 9 [F(1, 59)=2.466, P=.122], EMEQ 10 [F(1, 59)=.104, P=.746], EMEQ 11 [F(1, 59)=.093, P=.762], which shows that the mean for each item across all three application areas are similar. This supports the hypothesis that the WEKIT solution can be used to create expert models independent of the domain.

The results of the study show similar pattern across all three domains. For example, the median of EMEQ 11 was between 4-5 with large disagreement, while participants in all application domains seemed to strongly agree for EMEQ 1. The variance for EMEQ 11 can be explained with the complexity of the sensor framework built into the application. It is up to the expert author of the learning activity to decide where and when to stream sensor data. It is well possible that the chosen task may not have required data updates. Moreover, the automated adaptation of the activity based on sensor values may also hide that this happens from sensor data. We deem it therefore likely not all participants paid attention to the ‘data updating‘ possibility.

Average results show, however, that the expert participants found the expert model created by the WEKIT application to be usable for training students.

5.3.4 Knowledge Assessment

The aim of the Knowledge Assessment test was to evaluate the student participants performance after the training. The test was designed by the experts at the domain and almost each knowledge test question is testing knowledge acquired during consequent procedure step. In total there were nine procedure steps and 14 knowledge test questions in Medical domain and 15 procedure steps and 15 knowledge test questions in Aeronautics and Space training domain.

In the Aeronautics domain, there were 59 students in the experimental group, which used the player and 16 people in the control group which used paper based instructions.The group which used the application completed 66% of the questions correctly while the control group completed 63% of the questions correctly. The results (Z-score = 0.37 and p-value = 0.7) show there is no statistically significant difference between the two groups in Aeronautics domain.

In the medical domain, 73 students in experimental group used the player and 12 students were part of the Control group who used paper based instructions. The experimental group group completed 66% of the questions correctly while the control group completed 92% of the questions correctly. The results show there is no statistically significant difference between the two groups(Z-score = -1.7 and p-value

= 0.08).

In the Astronaut domain, 147 students in the experimental group used the player and 30 students were part of the Control group who used paper-based instructions. The experimental group group completed 66% of the questions correctly while the control group completed 63% of the questions correctly. The results show that there is no statistically significant difference between the two groups(Z-score = 0.3 and p-value = 0.76).

5.4 Conclusion

This study evaluated the validity and utility of expert models captured using the WEKIT solution in three independent test-beds. Results show that the WEKIT solution was rated positively in all three application domains with no statistically significant difference between test-beds. Experts agree that the model captured with the solution (and its affordances) are fit to be used for training in all three domains. The WEKIT solution implements the ID4AR framework (Limbu et al., 2018b) and all three models were captured using it. Therefore, the results of this study suggest the framework can be used more broadly across different domains for designing AR and sensor-based solutions for training. Moreover, the results of the knowledge assessment show that the AR and senor-based training is equally effective as the learning of the control group and there are positive effects with regards to acceptance (see (Guest et al., 2018)) and user experience (Xue et al., 2019). The use of the solution did not impede learning in comparison to the traditional methods and both group scored similarly in these knowledge assessment tests.

The WEKIT solution is a reference implementation of the ID4AR framework, an abstract framework for building sensor-based and AR based training applications. The presented evaluation results hold across the independent test-beds and thus support the claim that the framework can be used independent of application domain. The implementation and its evaluation underline that sensor-based AR systems are high-potential training tools. Moreover, they suggest that the adoption of the framework for designing AR training applications potentially can help mitigate risk, cost, and facilitate overcoming the complexity associated with their design and development.

5.4.1 Limitations and future work

Expert participants who peer evaluated the WEKIT solution based model did not have any pre/post sessions to help them prepare for the evaluation. The experts needed to recall their sessions to respond to the EMEQ questionnaire which may have affected the quality of the response. While the model was peer-evaluated by the other experts, there was no review of the model from the student's perspective. The knowledge assessment results in individual domains show none to very little significant difference in the learning performance of students who used the application than those who didn't. However, the assessment didn't take pre-knowledge and other factors into account. In addition, more work needs to be done to reap the benefits of the affordances of modern technologies such as AR to enhance the learning outcomes from the students. The WEKIT solution was a single solution to all three domains which was essential to meet the time and resource constraint.

Using the ID4AR framework to design specific solutions for individual domain can increase the affordances making it a more effective modelling tool.

Eventually, the work done so far has presented potentials and many opportunities for further development and research. Even though several milestones have been met in the development of the ID4AR framework, limitations exist. The framework itself is designed to be a support for training where experts are limited. The solutions designed with the framework are not for substituting the expert but for complementing them. While implementing the framework, the need to perform an extensive task analysis to select the proper set of IDMs on the domain still exists and is resource-intensive. In addition, with the evolving technology, the framework's pool of IDMs must expand to support the affordances of new technologies. The framework also does not claim explicating expertise and any tacit knowledge from the expert. While explicating the tacit knowledge is possible by rigorous manual means, by nature it cannot be done unobtrusively. Instead, the framework leverages on the performance metrics of the expert and visible attributes of expert performance to support training efficiently. While feedback is integral part of the framework in order to support training, the WEKIT solution has only focused on didactic methods and guidelines. No summative/formative feedback was provided based on expert data. Providing such feedback, especially formative, requires further research on both technology and methodology to be able to compare streaming data and experts recorded data in the physical time and space.(Schneider et al., 2018) and (Di Mitri et al., 2019) has been making significant efforts for achieving this feat. Their work so far has involved synchronised multi-modal data collection and annotation of such data which are crucial steps for being able to provide real-time feedback with sensor data.

Chapter 6

Calligraphy trainer: Assessing mental effort

Heard that the calligraphy artist passed out...apparently, he had one nasty stroke.

As shown in the study in chapter 5, the expert model recorded with WEKIT.One was evaluated positively for its usability in training. However, studies reported so far in the thesis do not specify how feedback should be provided using the expert model with out burdening students. To address this, the Calligraphy trainer prototype was built using the Instructional design for augmented reality (ID4AR) framework. Calligraphy trainer provides feedback to students using various modalities. The study investigates the mental effort imposed by feedback provided by the calligraphy trainer.

This chapter is published as: Limbu, B. H., Jarodzka, H., Klemke, R., and Specht, M. (2019c). Can you ink while you blink? Assessing mental effort in a sensor-based calligraphy trainer. *Sensors (Basel, Switzerland)*, 19(14)

6.1 Introduction

Several authors, including Di Mitri et al. (2018) and Specht et al. (2019) have elaborated on the reasons why sensors and multi-modality in learning are drawing so much attention. Using multi-modal data for training can have a significant impact on how learners learn (Schneider et al., 2018). Multi-modality refers to the communication and interaction practices in terms of multiple modes such as the textual, spatial and visual modes, where the use of several modes creates a single artefact or a message. Sensors can unobtrusively measure observable properties, which is ideal for capturing expert's performance as multi-modal data. Sensors can also monitor learner behaviour to provide feedback for effective learning using the captured expert performance and consequently, are capable of supporting deliberate practice. Deliberate practice is crucial in tedious tasks such as handwriting where a high amount of repetition is required to improve, and therefore, the practice should focus on improving a particular aspect of task (Ericsson et al., 1993). However, practising deliberately also requires additional mental effort because the learner needs to be conscious of his/her performance (Rikers et al., 2004). Thus, continuous real-time feedback, along with summative feedback, is needed to practice deliberately (Ericsson et al., 1993) and therefore, instructional designers need to take into consideration any additional mental effort that their instructional design may impose.

Handwriting is a complex perceptual-motor skill that requires many hours of practice to master (Feder and Majnemer, 2007). Perceptual motor skills, such as hand-eye coordination, are abilities which enables interaction with the environment by combining motor skills and human senses. Performance in such skills requires constant feedback from the environment which is collected from the human senses. Similarly, handwriting learning depends on how efficiently feedback is processed by the learner (Danna and Velay, 2015). This requires consistent practice for a long time. However, merely practising does not account for improved performance. Practice should be deliberate, i.e., aimed at improving the skill (Ericsson et al., 2018), but learners do not engage in deliberate practice spontaneously (Ericsson et al., 2007). Experts as mentors support the deliberate practice by providing constant feedback and guidance, which requires one-to-one settings (Carey, 2014). However, experts are scarce, and they cannot provide enough attention to each learner.

Additionally, the expert only has access to the final static image of the handwriting to provide feedback which ignores informative and dynamic aspects of handwriting, such as pressure and tilt of the pen (Asselborn et al., 2018). Therefore, it is difficult for the experts to provide the informative feedback required for deliberate practice. Sensors can be used to support the deliberate practice in learners by capturing an expert's performance as multi-modal data, which can then be used to provide continuous informative feedback and guidance.

The application "calligraphy trainer" for handwriting practice was built to support deliberate practice in novice calligraphy learners. It was built using the "Instructional design for Augmented Reality" (ID4AR) framework from Limbu et al. (2018b), which uses multi-modal data from experts to provide guidance and feedback. The application is designed to complement and support the expert rather than replace him/her. It uses various sensors to record an expert's performance, which can be used for practice. This allows experts to rapidly create learning content and spend less time on guiding and providing feedback.

A detailed account of the primary and supplementary feedback that a learner receives while practising handwriting is given by Loup-Escande et al. (2017). Primary feedback is naturally present in writing, namely: visual and proprioceptive feedback from the hand, which provides the sense of the hand's motion or position. The processing of primary feedback in handwriting occurs naturally and imposes intrinsic load on the learner. Therefore, this intrinsic load is inherent to the learning task. However, Danna and Velay (2015) argue that practising with supplementary feedback will enhance handwriting learning in comparison to receiving only primary feedback. The authors also acknowledge that adding supplementary feedback can increase the mental effort required for practice. For example, adding supplementary real-time visual information to handwriting learning, where vision is already used to process primary feedback, can increase the mental effort for the learner.

Similarly, using haptic devices to provide additional feedback might result in an additional mental effort, as proprioceptive feedback naturally exists in handwriting learning. Loup-Escande et al. (2017) examined Danna and Velay (2015) suggestion to augment the strokes with supplementary information to provide additional visual feedback and found that such type of interventions does lead to additional mental effort. Similar phenomena can be observed with supplementing haptic feedback in a proprioceptive task. However, they did not explore the mental efforts imposed by auditory feedback. The auditory modality is not naturally found in handwriting, and it can be used to provide supplementary feedback without additional mental effort (Mayer and Moreno, 2003). However, auditory feedback has received little attention, mainly because of the difficulties inherent in providing easily understandable auditory feedback (Sigrist et al., 2013). Baur et al. (2009) reported significant improvements in the writing performance of people with Writer's cramp when the grip force was translated into auditory feedback. The calligraphy trainer implements the suggestions of Danna and Velay (2015) and Loup-Escande et al. (2017) by augmenting supplementary visual and auditory feedback. Therefore, this study aims to evaluate the mental effort imposed by the calligraphy trainer and the types of feedback provided by the application. As such, we examine the following research questions, using the calligraphy trainer.

- Is the System Usability Scale (SUS) score of the prototype at an acceptable level (above 68) (Brooke et al., 1996)? If not, does it co-relate of mental effort?
- Is the mental effort imposed on the treatment group by the feedback mechanism including auditory feedback significantly higher/lower than the control group's mental effort?

6.2 Background

6.2.1 Use Case Description

Handwriting relies on fine motor movements of the hand to create unique styles of writing. The fundamental aspect of handwriting is to control the pressure applied to manipulate the thickness of the strokes and to glide the pen in the correct path. Common mistakes found in beginners include quickly forgetting to remind themselves to maintain the basic factors such as grip force, posture, and angle of the pen (Thorpe, 2013). Besides, they quickly lose patience, which leads to quickly drawn strokes rather than slow, steady ones. Therefore, constant feedback from the expert is crucial to ensure deliberate practice as beginners are unable to monitor themselves. The calligraphy trainer used in this study is built using the ID4AR framework which provides continuous feedback to the learners in order to assist them to practice deliberately. The framework is briefly introduced in the following section.

6.2.2 ID4AR Framework

The ID4AR framework proposed Limbu et al. (2018b) supports instructional designers to design multi-modal systems with augmented reality and sensors for supporting deliberate practice. The framework exploits sensors' capabilities to record performance data for training. It is designed to be domain-independent (Limbu et al., 2019b) and is built in close collaboration with experts in three different domains. The framework's motivation to capture expert model independent of domain-specific implementations was evaluated in Limbu et al. (2019b). To do this, the "WEKIT" application, built using the ID4AR framework, was used. Before evaluating the framework itself, this application was evaluated in terms of having met the framework requirements, by the experts from the three domains, including experts who helped design the framework (Limbu et al., 2018c). Then, the ID4AR framework was evaluated by capturing an expert model with the help of an expert who was familiar with the application. This model, which underwent no further post-processing, was then used by the other experts and rated to meet the training requirements. Results in Limbu et al. (2019b) showed that the framework can be used across various domains. Below, we provide details on how "calligraphy trainer" was designed using the ID4AR framework.

6.2.3 Prototype Description

To implement the ID4AR framework, the calligraphy trainer implements Instructional design methods (IDMs) from each component of the model depending on the identified attributes of calligraphy (see Table 6.1). Attributes are characteristics of writers or the process of writing that influences the outcome of handwriting. Two categories of attributes were identified, which are 1. Non-expert based, and 2. expert-based. Non-expert rules are fundamental, universal rules of thumbs that do not require experts to generate feedback and are prioritised when generating feedback. On the other hand, feedback based on an expert's data are parameters that are recorded from the expert using sensors. These parameters are influenced by the context of practice, e.g., the style and the character which the expert demonstrates during recording. The types of IDMs implemented for each of the attributes are detailed in Table 6.2. The IDM *Augmented paths* for "learning task" displays the character which the expert recorded. Learners trace over these for practice. The IDM *Haptic feedback*, *Object enrichment* and *Auditory feedback* are implemented to provide feedback on procedural information while the IDM *Animation* provides supportive information such as speed and path, on the learning task. *Summative feedback* is provided by collecting, visualising, and comparing learner's data with the expert's data by using the Visual inspection tool (Di Mitri et al., 2019). More details on the implementation of these IDMs are provided in the following sections.

Table 6.1: Types of expert attributes identified.

Non-Expert Based	**Expert Based**
1. Force used to grip the pen	1. Pressure used to create the strokes
2. Angle at which the pen is held	2. Similarity of the stroke structure
3. Body posture	3. Speed of writing

Table 6.2: Mapping of attributes with IDMs in Calligraphy Trainer.

Attributes	**IDMs**	**Implementation**
Learning Task		
Alphabets Structure	Augmented Paths	Displayed on tablet for tracing or imitating, color of the stroke changes when the color stroke is out of bounds
Procedural Information		
Force used to grip the pen	Haptic feedback	Vibrate myo when the grip is too tight or the angle is beyond the threshold
Pressure used to create the strokes	Object enrichment	Stroke thickness is directly proportional to the pressure, The stroke darkness/lightness is also directly proportional to the pressure
Supportive information		
Speed of writing, alphabet structure	Animation	animation depicting the speed and the path in which the alphabet was written
Part task practice		
Over all performance	Summative feedback	Summative results produced by comparing with the expert recording

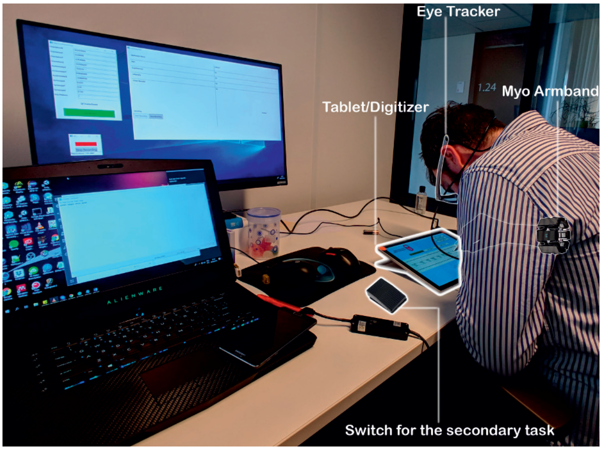


Figure 6.1: System Model for supporting the framework.

Hardware Description

The hardware setup consists of a Microsoft™Surface Pro Tablet, the Surface Pen and a Myo™Armband. The Surface Pen and the Myo™armband both act as an input device and feedback systems. The Myo™armband consists of EMG sensors (electromyography) that reads muscle activity and also reads hand gestures and orientation with the embedded accelerometer and gyroscope. It also includes a vibration motor to provide haptic feedback. The capacitive Surface Pen and the digitizer on the Surface tablet act as the main canvas for the learner to practice handwriting. The pen and the digitizer together can read the pressure applied while creating the stroke and the angle at which the pen is held, normal to the digitizer surface. The tablet also runs the multi-modal Learning Hub application (Schneider et al., 2018), which synchronises sensor data and acts as a gateway for sensors to communicate as well. The calligraphy trainer records performance data with these sensors and also, provides the users with real-time feedback during practice using the captured expert's data. Figure 6.1 depicts whole setup used in the study.

Software Description

The system consists of two main components: the recorder to record the expert's performance and player for training learners based on the expert's performance (see Figure 6.2). The recorder records all the data needed for the learner to perform the task. In the recorder, values for identified attributes of calligraphy are captured from the experts (see Table 6.1). A separate process that collects data from the Myo™armband runs separately in the background from the main application, which is the calligraphy trainer. The multi-modal learning hub (Schneider et al., 2018) is used to collect synchronised data from the Myo™armband and the stylus pen. The player loads the data for practice. It provides guidance and feedback using the recorded data by comparing learner's current attribute values to the expert's values in real-time. It also stores learner's performance for summative feedback, which can be used both by the learner and the expert for reflection.

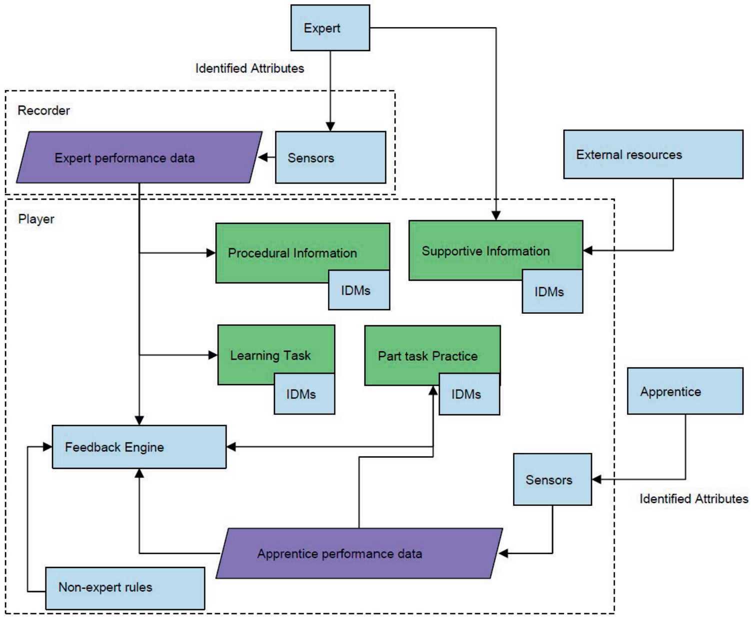
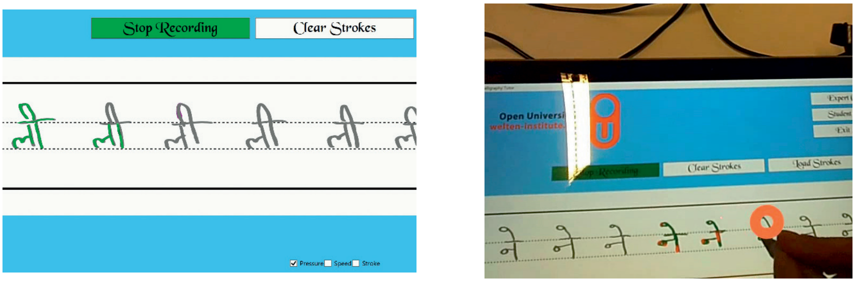


Figure 6.2: System Model for supporting the framework.

Calligraphy Trainer

The calligraphy trainer supports two different roles, for the experts and for the learners. The calligraphy trainer allows experts to draw strokes which are saved as data into an Ink Serialised Format (ISF) file and the sensor data that is stored as json files. On the other hand, the learners can load the data that was saved by the expert to practice. As shown in Figure 6.2, non-expert based attributes are hard-coded into the feedback engine. For the expert based attributes, the application provides feedback by referencing the expert's data as the learner practices. Feedback is provided for three expert based attributes that the learner can choose to turn on or off (see Table 6.2). The supplementary feedback for the pressure applied is given by varying the saturation of the colour (see Figure 6.3a). When the pressure is above the expert's pressure, the colour gets darker, and when the pressure is below the expert's pressure, the colour starts to get lighter. However, the primary feedback for pressure which is given as the thickness of the stroke, is always present. Similarly, feedback on the stroke structure is given by changing the colour. When the learner's stroke goes out of bounds from the expert's stroke, the colour of the stroke changes to red (see Figure 6.3b). The feedback on the speed of the stroke is auditory. Learners hear a buzzing sound when they are over the speed of the expert. No auditory feedback is given when the learners are below the expert's speed. Only one non-expert based attribute is implemented for feedback. Feedback for the force used to grip the pen is implemented using Myo$^{\mathrm{TM}}$, which provides haptic feedback when the user holds the pen too tightly.

In addition to the feedback, guidance on the process to write the character was provided using the IDM *"Augmented Paths"* by displaying the character drawn by the



(a) Pressure feedback with saturation.

(b) Stroke feedback with colour.

Figure 6.3: Feedback provided by the calligraphy trainer.

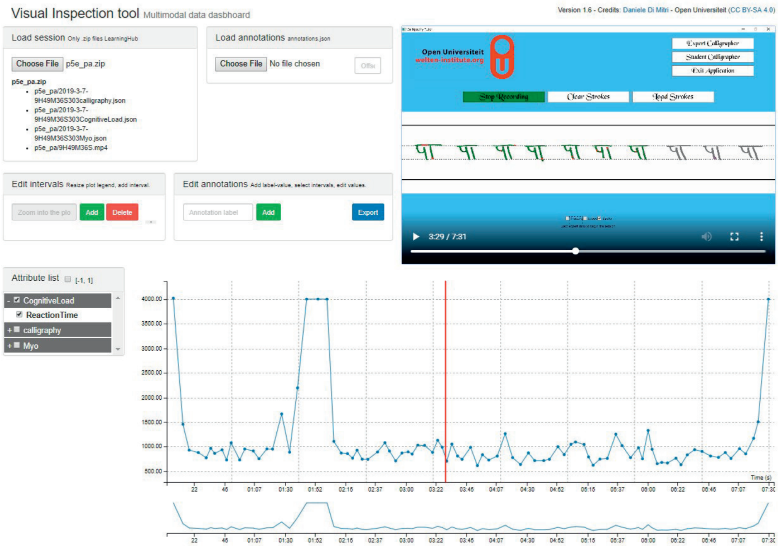


Figure 6.4: Visual Inspection tool for providing summative feedback.

expert as a semi-transparent image. Further supportive information on the character's speed and the sequence was provided using an animation. The semi-transparent character was overlapped with a running animation which played according to how the expert drew the character. This guided learners on how the pen is moved, which is of more importance than the shape of the character itself (Jarman, 1979; Morikawa et al., 2018). The IDM *summative feedback* was provided by the expert with the help of the recorded data using the Visual Inspection tool (Di Mitri et al., 2018) (see Figure 6.4). The application records temporal data with all the sensors which can be loaded in the Visual Inspection tool along with a video recording of the performance.

6.3 Methods

In order to evaluate our research questions considering the mental effort of participants evoked by different types of feedback, we designed a formative study. While learners had to write characters based on the given expert model they received, either real-time feedback in multiple modalities or they did not receive feedback. Additionally, we measured the usability of the system to avoid an effect of usability issues on the participant's mental effort.

6.3.1 Participants

The study was conducted with ten randomly selected PhD students working in the educational science and technology department at the Open University of The Netherlands. Out of the 10 participants, six were female, and 4 were male. All the participants were right-handed. None of the participants had any experience writing the script used in the study. Participation was completely voluntary.

6.3.2 Apparatus

The apparatus for the study consisted of the calligraphy trainer, which is the main application for the users. It runs on the surface tablet and provides data for stoke pressure and angle, with the help of the pen. It also displays the ink stroke and provides visual and auditory feedback on the tablet. The experts can record data, and the learners can practice with the help of the recorded data using the calligraphy trainer. It also guides learners on how to draw the character using the expert's data. The Myo$^{\mathrm{TM}}$armband is used to provide haptic feedback to the learner. The armband uses an electromyogram to detect the tension in muscles, which co-relates to how hard the learner is gripping the pen.

Additionally, the application for recording the reaction time of the participants with a USB switch was also used. It recorded the time participants took to react to the auditory stimuli of the secondary task in milliseconds. The eye tracker glasses from SMI$^{\mathrm{TM}}$were used during the study to collect eye-tracking data. The eye-tracking data was used to measure the mental effort using the pupil dilation and can also help gain further insights into the software usability if needed.

6.3.3 Procedure

Before beginning, participants were informed about the study and were asked to sign the informed consent. Then, the participants were briefed on the task they needed to perform. In this briefing, they were informed that they were expected to replicate an expert's writing. During this step, the participants in the treatment group were also briefed on the type of feedback they will be receiving. After this, the sensors were calibrated, and the participants were allowed to freely practice using the stylus until they felt comfortable using it (in a different drawing application). When the participants said they were ready, the study began by loading the first character. Participants in both groups performed four iterations for each character. The treatment group received feedback during this while the control group did not. Participants in both groups were asked to fill in the questionnaire for the mental effort (see Mental effort in Materials and measures) at the end of each iteration, therefore, 12 times during the whole study. The participants in both groups also performed the secondary task during the study. At the end of the study, participants were asked to fill in the SUS questionnaire. They were given opportunities for open comments on the calligraphy trainer and were thanked for their participation.

6.3.4 Materials and Measures

Usability

The System Usability Scale (SUS) was used to measure the usability of the application. SUS is an industry-standard tool for measuring system usability, which refers to the ease of use of an application. It consists of a 10 item questionnaire with five response options for participants, ranging from “strongly agree” to “strongly disagree”. The SUS scores calculated from individual questionnaires represent the system usability. Scores for individual items on the SUS are not meaningful on their own. SUS yields a single number between 0 to 100, which represents a composite measure of the overall usability of the application. The acceptable SUS score is about 70. SUS is an easy scale to administer and can be used on small sample sizes with reliable results. It can effectively differentiate between usable and unusable systems (Brooke, 2013). While SUS is not a diagnostic tool, further usability analysis can be done with eye-tracking data if required. Our aim behind using the SUS was only to confirm that the obtained results on mental effort were not influenced by usability issues of the application.

Mental Effort

Beginner calligraphers need to continually monitor themselves to practice deliberately, for which constant feedback from the expert is crucial. Monitoring their performance while practising is cognitively demanding. Therefore, the application should not levy extraneous mental effort, which is a negative load caused by ineffective instruction (Brunken et al., 2003). To keep the mental effort to a minimum during practice, we adopted Danna and Velay (2015) proposed solutions for adding supplementary visual feedback. They suggested that the kinetic variables of the movement should be represented in the stroke itself, and summative feedback should be introduced after, and not during, the execution of the gesture. The calligraphy trainer provides feedback for the kinetic variables such as speed and pressure during the execution of the stroke. No complex feedback is provided and the learning task is simply to reproduce the stroke. Contrary to Frenoy et al. (2016) implementation of the system, the calligraphy trainer relies on the expert and the expert’s data for providing feedback. While summative feedback is provided at the end of the practice session, this was not relevant for this study.

The mental effort was measured using dual-task methodology (Brunken et al., 2003). Dual-task methodology requires participants to perform a secondary task in parallel to the primary task. The secondary task in this study required the participants to react to auditory stimuli (a gong sound) by pressing a switch as soon as they could with their non-dominant hand. The stimuli were presented at random intervals between two to six seconds. The time required by the participants to react was recorded. Lower reaction time denotes lower mental effort due to free working memory available for processing the secondary task.

Table 6.3: SUS scores.

Groups	Average SUS Score
Control Group	78
Treatment Group	87.5
Combined	82.75

The participants also wore an eye tracker during the study. The eye tracker records various types of data, such as gaze positions, pupil dilation, saccade rate, fixations, and blink rates. Data types such as pupil dilation, saccade rate, and blink rates are co-related to the mental effort (Holmqvist et al., 2011). Additionally, Paas et al. (2008) subjective rating scale for mental effort (will be referred to as mental effort questionnaire) was also used to complement the collected data on mental effort. Participants filled in the questionnaire after each iteration for all characters by selecting a response between 1 (very, very low mental effort) to 9(very, very high mental effort).

6.3.5 Design

Participants were randomly assigned to the treatment and the control group. In the control group, the participants used the same setup, but the feedback was not given. Participants in both the group practised each character, “Ne”, “Pa” and “Li” in the presented order, in four iterations with each iteration requiring the participants to write the character ten times. All the participants reacted to the secondary task during the whole duration of the study and responded to the mental effort questionnaire at the end of each iteration. The treatment group followed the same procedure but received feedback on the kinetic variables. For each character, the first three iterations were performed with feedback on one kinetic variable while the last iteration was performed with feedback on all three of them. However, the order of the first three iterations for individual participants was assigned following the Latin square design to ensure that all participants did not go through the same sequence of kinetic variables.

6.4 Results

6.4.1 SUS Scores

A paper-based SUS questionnaire was administered at the end of the study for the participants in both groups. The SUS score for the Control group (78) and the treatment group (87.5) is at an acceptable range (see Table 6.3). Both groups had an equal number of participants (N = 5). We conducted the Shapiro-wilk test on the SUS items, which showed that none of them were normally distributed. There is statistically no significant difference in SUS scores based on the group, F (7, 2) = 16.943, $p = 0.057$. The SUS score for both groups together (82.75) is at an acceptable range as well.

Figure 6.5: Mean of Self-reported mental effort between two groups.

6.4.2 Mental Effort

Self Reported Mental Effort

Self-reported questionnaires were used to collect the response on the mental effort required during each iteration. The mean response for both the control and treatment group according to the type of feedback is presented in Figure 6.5. We conducted the Shapiro-wilk test, which showed that control group iteration Ne_Pressure, Ne_all, Pa_pressure, Pa_stroke, and Pa_all were not normally distributed. While in the treatment group, iterations Ne_All and Pa_stroke were not normally distributed. A Manova was conducted to compare the mental effort between the control and the treatment group. There was no significant difference in the self-reported mental effort for all the iterations between the control and treatment group (see Figure 6.5).

There is statistically no significant difference between the groups for the reported mental effort in Pressure: F (3,4) = 1.436, p = 0.357, Speed: F (3,4) = 0.987, p = 0.996 and Stroke: F (3,4) = 0.017, p = 0.730. There is also statistically no significant difference between the groups for reported mental effort in combined feedback scores F (3,4) = 0.017, p = 0.514. There is little to no evidence that the self report data provides for effect of the treatment on the mental effort of the user.

Reaction Time on Secondary Task

The secondary task logged the participant's reaction time in milliseconds (see Mental effort in the Methods section). Any data point lower than 250 ms and more than 3750 ms was removed to account for accidental presses. Then, the reaction time was transformed into Log^{10}. The mean reaction time for all the iterations between the control and treatment group is presented in Figure 6.6.

We conducted the Shapiro-wilk test, which showed that in control group sessions, Ne_Speed, Ne_Pressure, Ne_All, Pa_Stroke, Pa_Speed, Pa_All, Li_Stroke, Li_Pressure, and Li_All were not normally distributed. While in the treatment group sessions,

Chapter 6

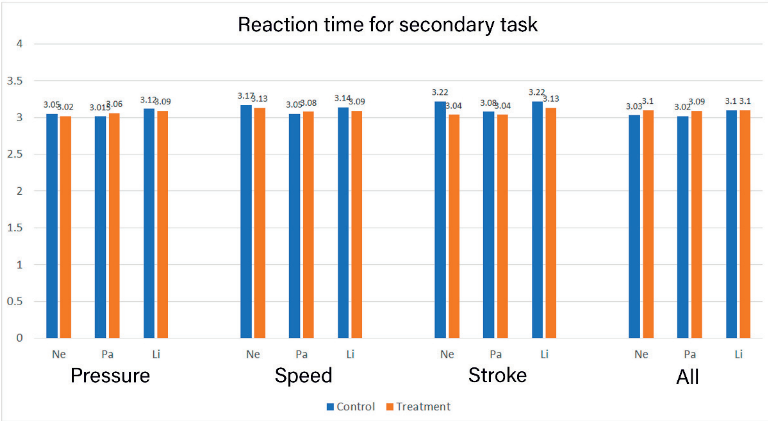


Figure 6.6: Mean of Reaction time between two groups [in Seconds].

Ne_Stroke, Ne_Pressure, Ne_All, Li_Stroke, Li_Speed, Li_Pressure, and Li_All were not normally distributed. There is statistically no significant difference between the two groups in reaction time for Pressure: F (3,70) = 1.908, p = 0.136, Speed: F (3,87) = 1.439, p = 0.237 based on the group. However, there is a statistically significant difference between the two groups in reaction time for Stroke: F (3,89) = 7.672, p = 0.000, scores based on the group. The effect for the group yielded an F ratio of F (1,91) = 22.848, p = 0.000 for character "Ne" and an F ratio of F (1,91) = 5.485, p = 0.021 for "Li" indicating significant difference between control and the treatment group for character Ne and Li for the Stroke feedback while there was no significant difference between the groups in character "Pa". There was also statistically no significant difference in reported reaction time for combined feedback scores based on the group, F (3,95) = 2.653, p = 0.051.

Time Taken

The mean time taken in seconds to complete each iteration by the groups is presented in Figure 6.7. The treatment group took a longer duration to complete the task as compared to the task in all iterations in comparison to the control group.

A Shapiro-Wilk test on the variables showed that all the data for time taken was normally distributed in both the groups. We conducted a Manova to compare the means between the two groups. There was a statistically significant difference between the two groups for mean time taken to complete the task for Pressure: F (3,6) = 6.378, p = 0.027, Speed: F (3,6) = 10.683, p = 0.008 and Stroke: F (3,6) = 7.628, p = 0.018. The effect for the group yielded an F ratio of F (1,8) = 15.971, p = 0.004 for character "Pa" and an F ratio of F (1,8) = 24.031, p = 0.001 for "Li" indicating significant difference between control and the treatment group for character Pa and Li while there was no significant difference between the groups in character "Ne" while providing pressure feedback. The effect for the group yielded an F ratio of F (1,8) = 22.800, p = 0.001 for character "Ne", an F ratio of F (1,8) = 24.61, p = 0.001 for "Pa" and an F ratio of F (1,8) = 20.147, p = 0.002 for "Li" indicating a significant difference between control and the treatment group while providing speed feedback. The effect for the group while providing stroke

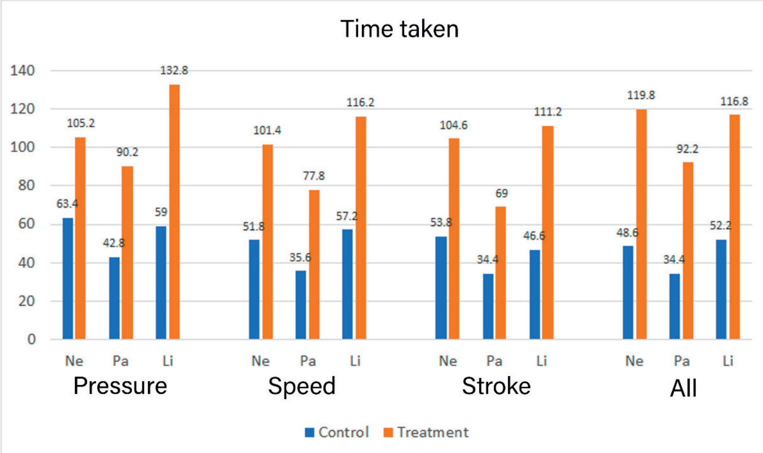


Figure 6.7: Time taken by the two groups [in Seconds].

based feedback yielded an F ratio of F (1,8) = 6.849, p = 0.031 for character "Ne", an F ratio of F (1,8) = 14.025, p = 0.006 for "Pa" and an F ratio of F (1,8) = 25.920, p = 0.001 for "Li" indicating significant difference between control and the treatment group.

There was also a statistically significant difference in time taken for combined feedback scores based on the group, F (3,6) = 23.178, p = 0.001. The effect for the group while providing stroke based feedback yielded an F ratio of F (1,8) = 7.204, p = 0.028 for character "Ne", an F ratio of F (1,8) = 38.501, p = 0.000 for "Pa" and an F ratio of F (1,8) = 36.200, p = 0.000 for "Li" indicating significant difference between control and the treatment group for all characters.

6.4.3 Eye Tracker

The head-mounted eye tracker was used to collect eye-tracking data. We used the pupil dilation from the eye tracker data for measuring the mental effort. The pupil dilation is found to be directly proportional to the mental effort (Szulewski et al., 2015). Figure 6.8 shows a larger pupil diameter in the treatment group, which signifies greater pupil dilation and thus, more mental effort in the treatment group.

A Kolmogorov-Smirnov test indicates that the pupil dilation for the right eye in all intervention does not follow a normal distribution. A Manova test showed that based on the group, there was statistically significant difference in Pupil diameters for Stroke: F (6,26482) = 450.713, p = 0.000, Speed: F (6,24907) = 1593.861, p = 0.000 and Pressure: F (6,22071) = 2274.819, p = 0.000. There was also a statistically significant difference in pupil diameters for combined feedback based on the group, F (6, 22552) = 644.641, p = 0.000. The pupil diameter is provided only for the right eye in Figure 6.8, as one eye, is enough to estimate the mental effort.

The effect for the group on the diameter of the right eye pupil, while providing pressure based feedback, yielded an F ratio of F (1,22076) = 10616.929, p = 0.000 for character "Ne", an F ratio of F (1,22076) = 4751.480, p = 0.000 for "Pa" and an F ratio of F (1,22076) = 6295.214, p = 0.000 for "Li" indicating significant

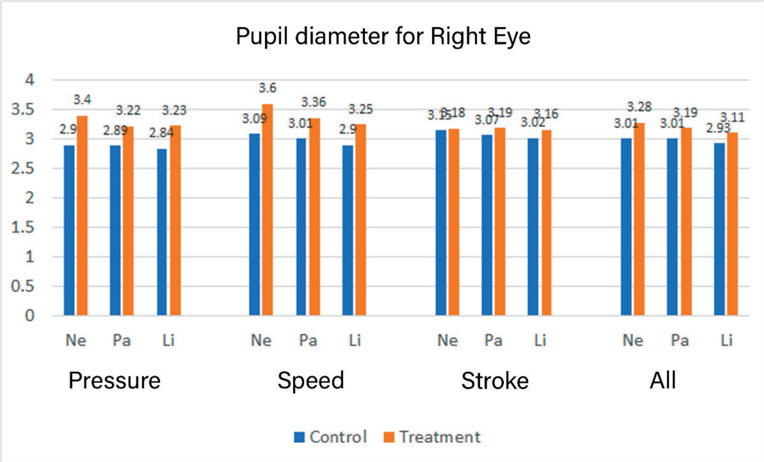


Figure 6.8: Pupil diameter [in millimeters].

difference between control and the treatment group for all characters. The effect for the group on the diameter of the right eye pupil, while providing stroke based feedback, yielded an F ratio of F (1,26487) = 49.061, p = 0.000 for character "Ne", an F ratio of F (1,26487) = 575.350, p = 0.000 for "Pa" and an F ratio of F (1,26487) = 707.752, p = 0.000 for "Li" indicating significant difference between control and the treatment group for all characters. Similarly, the effect for the group on the diameter of the right eye pupil while providing speed based feedback yielded an F ratio of F (1,24912) = 7062.894, p = 0.000 for character "Ne", an F ratio of F (1,24912) = 4382.158, p = 0.000 for "Pa" and an F ratio of F (1,24192) = 4784.584, p = 0.000 for "Li" indicating significant difference between control and the treatment group for all characters. The effect for the group on the diameter of the right eye pupil, while providing all types of feedback, yielded an F ratio of F (1,22557) = 2321.941, p = 0.000 for character "Ne", an F ratio of F (1,22557) = 1116.390, p = 0.000 for "Pa" and an F ratio of F (1,22557) = 1274.489, p = 0.000 for "Li" indicating significant difference between control and the treatment group for all characters.

6.5 Discussion

This paper presents a formative pilot study to evaluate the calligraphy trainer application considering the mental effort involved in using the application with different types of provided feedback. The tool supports deliberate practice in novice calligraphy learners (Limbu et al., 2018b) and assists the experts to create learning content quickly. In addition, it provides feedback and guidance based on expert data while recording the learners' performance, which allows reflection by the expert on the learning process itself. The expert also decides the content along with the type of feedback, based on the task parameters that the learner needs to train on. This is different from the approach of Frenoy et al. (2016), who developed a model for providing the correct feedback type based on sensor data. The calligraphy trainer was designed such that the identified learning parameter can be isolated and trained individually until mastered before practising more complex scenarios. As such, only the feedback on a single parameter is provided at a time, unless chosen not to do so by the expert. The calligraphy trainer application provides two types of supplementary visual feedback, which are integrated into the stroke of the pen and an auditory feedback. By evaluating the mental effort required to process the feedback individually and also, when combined, the design of the feedback can be improved.

During the study, participants were required to load the expert data and write the characters. The study took 30 min to 1 h, depending on the group because of the necessity for manually segregating the data into proper sessions to avoid extremely long temporal data. The participant spent most of the time waiting for the data to be logged and saved as this was done manually by the examiner. Future versions of the application are expected to handle this automatically in the background. At the end of the study, the SUS questionnaire was used to evaluate the overall usability of the application. The main objective of this study was to test the mental effort imposed by the feedback. However, we consider it essential to confirm that the obtained results were not influenced by usability issues of the application, which was explored by the first research question. Therefore, to confirm this, participants filled in a SUS questionnaire at the end of the test. Scores from the SUS show that the application is well over the acceptance level; therefore, we assume that the usability of the application was not a determinant factor in the observed results about the mental effort imposed by the feedback.

To answer the second research question, we measured the mental effort imposed by the type of feedback with self-reports, dual-task methodology, and pupil dilation from the eye tracker. The results from the self-reported mental effort show that the treatment group reported higher mental effort in all three characters when compared to the control group, only when all three types of feedback were provided simultaneously. Nevertheless, both groups reported the mean mental effort for each type of feedback to be 5 or higher. This may signify that handwriting learning requires naturally higher mental effort (Feder and Majnemer, 2007), and instructional designers should design their feedback keeping this in mind. Similarly, the results from the reaction time show an identical pattern to the self-reported mental effort. Only the combined feedback had consistently higher mental effort across all three characters. However, the reaction between the two groups was nearly identical across all interventions. The mean reaction time was above 3 seconds in all interventions with 4 seconds being the maximum. This was higher than we expected but is in line with the argument that learning handwriting requires a high mental effort. The individual feedback interventions had mixed results in self-report and reaction time, with some characters showing higher mental effort in the control group (see Figures 6.5 and 6.6). This contradicts our assumption for individually provided feedback, where we expected the base mental effort to be similar in both the group. If any deviations, the treatment group was expected to have higher mental effort due to the requirement for processing the additional supplementary feedback. It should also be noted that the time taken by the treatment group to finish the task was significantly longer than the control group. Perceiving and processing supplementary visual feedback requires additional time to be compatible with the immediate corrections required during handwriting (Danna and Velay, 2015). The time taken may have had contributed to the results in reaction time in the treatment group.

The results collected from the eye-tracker on pupil dilation show consistently higher mental effort for the treatment group across all interventions and across all characters in each intervention. This is not in line with the results from the self-report and the reaction time for individual feedback intervention. While the mental effort on individually given feedback is inconclusive, the participants in the treatment group have always reported higher mental effort in all three metrics, namely, self-report, reaction time and pupil dilation for combined feedback intervention. This supports the calligraphy trainer's approach to isolate individual parameters for practice rather than the approach of Frenoy et al. (2016) to have a model decide the feedback to be given. This uncertainty of the feedback the learner is going to receive next might add overhead costs to process them. On the other hand, the approach used by the calligraphy trainer reduces the complexity of processing the feedback for practising by isolating a single parameter and the feedback on the parameter.

The results on mental effort for speed, which was given by the auditory channel was not conclusive in terms of requiring lower mental effort in comparison to pressure and stroke. The speed feedback did not have noticeably lower or higher mental effort in the treatment group than the control group. In contrast to Mayer and Moreno (2003) suggestion, it is unclear if using auditory modality results in lower cognitive load in this case. Our implementation of the auditory feedback consisted of a simple buzz sound when the learner went over the expert's speed in that particular stroke. The auditory feedback was kept simple to keep the processing cost of minimal, but this might have resulted in the break of flow for the learners when suddenly interrupted by the buzzing sound. A similar pattern was seen in Loup-Escande et al. (2017) finding that the feedback for speed provided by producing large circles on top of the stroke resulted in higher mental effort. These circles break the natural flow that the user is in during the writing process. Auditory modality has also been used to provide feedback on the grip force by converting the EMG data into sounds to assist learners to control their grip force (Baur et al., 2009). In contrast, this study provided feedback on the grip force by haptic means. The haptic feedback was provided with the Myo™armband, but it was not evaluated in this paper and was given to all the participants in both the group. Proper ways to provide supplementary haptic feedback for a proprioceptive task, where motor modality is already being used is unclear and lacks research (Danna and Velay, 2015), unlike the visual modality.

6.6 Conclusions

This pilot study is a formative study aimed at evaluating the mental effort imposed by the supplementary feedback provided by different versions of the calligraphy trainer. The calligraphy trainer leverages on the recent advancement of sensor technology and digitizers, to explore supplementary real-time feedback on the writing process. The high SUS score enabled us to ensure that the usability was not a factor in determining the mental effort. Except for self-reports, results from the reaction time and pupil dilation show that the mental effort in the treatment group is only slightly higher. The effect of modality was also unclear from the results. The auditory feedback did not result in a comparatively lower mental effort as was assumed.

Observations from the eye-tracking data show that the learners were fixated on the tip of the pen during the whole process (see Figures 6.3b and 6.9) and,

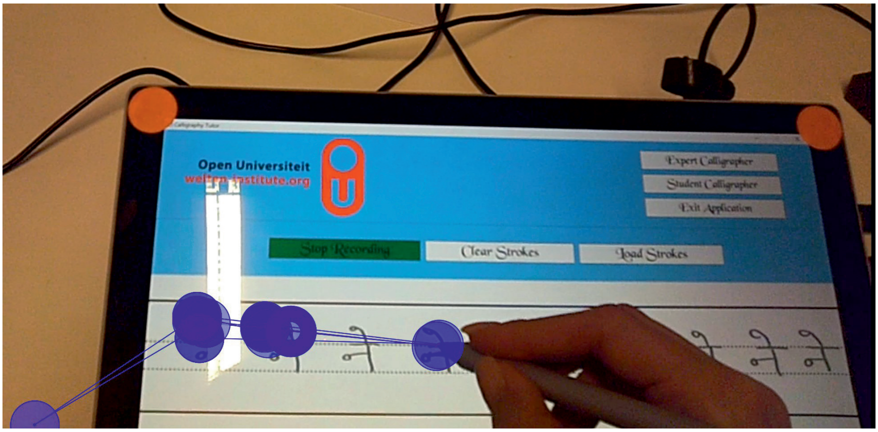


Figure 6.9: Visual scan path of the participant while writing.

therefore, feedback should be given immediately after the stroke is generated. Since handwriting learning for a novice is usually a high mental effort task, the best course of action for designers is to try to minimise the mental effort as much as possible. Danna and Velay (2015) recommend, supplementary feedback should be provided in a different modality. Only when required, supplementary visual feedback can be provided by augmenting the information on top of the stroke. Currently, meaningful versatile haptic feedback that can be used to convey different types of information is lacking. The most common implementation of such feedback is a basic vibration. The Myo^{TM}armband can alter the duration of vibration, but it is difficult to interpret this feedback in the context of calligraphy meaningfully. Similarly, audio-based feedback can provide detailed vocal feedback, but practising calligraphy requires quick adaptation to the feedback (Teulings and Schomaker, 1993). Therefore, auditory feedback should be designed to be quick and take minimal mental effort to process.

In conclusion, this formative pilot study indicates that calligraphy trainer's feedback does not impose excessively high mental effort on the user. However, the base mental effort from the control group without the feedback was still high. Further study is required to determine if this was the result of the supplementary haptic feedback given by the armband or an intrinsic load. Similarly, using auditory feedback did not result in lower mental effort. Even though the reported mental effort was similar to the other visual feedback, as shown by the pupil dilation and reaction time, the design of the auditory feedback must be improved to make better use of the modality. Furthermore, new methods for using other modalities instead of visual mode should be explored to reduce the overall load.

In addition to designing proper feedback to lower the mental effort in the learner, the learning process itself can be designed to lower the intrinsic mental effort required. The framework from Limbu et al. (2018b) recommends isolating the task parameters and practising them in order of incremental difficulty. The results of the mental effort clearly show that mental effort is higher when all the task parameters are practised together. Such scenarios should be practised at the end when all individual parameters have been mastered. Reducing the intrinsic mental effort in an individual practice session will allow learners to focus more on the feedback for the parameter being practised. This will lead to deliberate practice, which can result if efficient and effective learning.

6.7 Limitations

The study is limited by the number of participants. While ten participants are considered enough to study the usability of the application, it is difficult to generalise findings on the mental effort with just 5 participants in each group. However, this is a formative pilot study and is expected to be up-scaled, which may result in concrete conclusions. The study compares the mean of the types of feedback to the mean of the iteration in the control group. Doing so does not take into account the decrease in the mental effort in the control group due to repetitive practice. Besides, the haptic feedback which was provided to both groups might have affected the outcome of the mental effort. Similarly, the effect of feedback modalities on the mental effort could not be compared. The study also did not take into account the learning outcomes between the two groups. Providing feedback can induce additional mental effort, but they are crucial to learning and therefore, must be taken into account.

Chapter 7

Discussion

7.1 Summary and Overview of the Findings

This chapter begins with the overview of the main findings of the studies reported in this thesis and their relationships. Figure 7.1 provides an overview of the various studies done in the project and their relationship to the primary contribution of the project, namely the Instructional Design for Augmented Reality (ID4AR) framework. This chapter concludes with the discussion of the limitations and the future works of the project.

The research was conducted to find answers for the main research question "How can Augmented reality(AR) applications be designed for deliberate practice of complex skills?", which was divided further into the following sub questions below. These sub questions were addressed by conducting a design-based research, which is reported in the chapters of this thesis.

RQ1 Which design patterns can be used in AR to train different types of skills?

RQ2 How can design patterns be systematically implemented in AR to support deliberate practice?

RQ3 How can expert performance be modelled and evaluated?

RQ4 How can feedback be designed with-out imposing high mental effort on students?

7.1.1 [RQ1] Which design patterns can be used in AR to train different types of skills?

RQ1 was addressed by conducting the literature study reported in chapter 2. The literature study reviews 78 studies that implements AR and sensor-based technologies (referred to as only AR from here on) to explore how such technologies can be used to train complex skills. The abundant results from the preliminary search indicates the evident potential of AR as training platform. However, only prototypes that involved mentors in the training process were included for further review to keep mentors as a part of the AR-based training. The first phase of the literature study involved extracting instructional designs of the selected prototypes. This

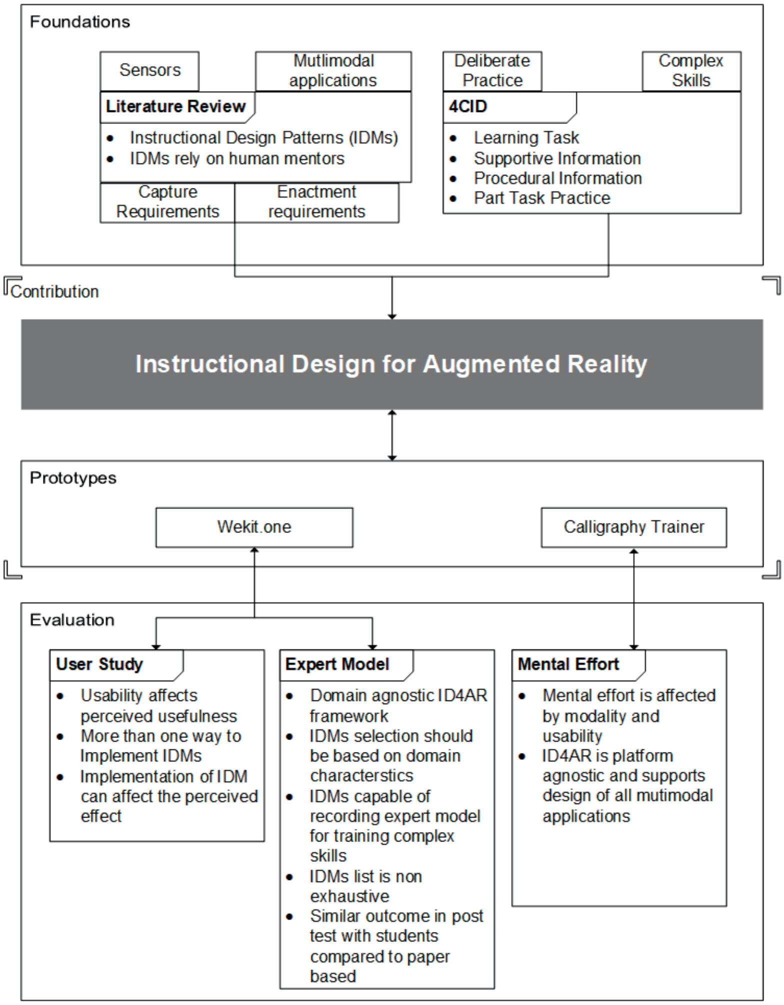


Figure 7.1: Findings and outcomes of the project

resulted in a non-exhaustive list of AR-based instructional design patterns (IDMs) which utilised human mentors for training complex skills. This list formed the basis of further works in forthcoming chapters, for example, generating the ID4AR framework and designing the prototypes.

All IDMs are defined with core functionalities, recording requirements and re-enacting requirements. Recording requirements defined how to record mentors performance while the re-enacting requirements define how to use the recorded performance for training. IDMs are abstract building blocks and therefore, can be used to design an AR-based training environment that relies on mentors for facilitating training. The IDMs are also tagged with the type of skill that is used to train, cognitive skill for example, in order to help the instructional designers choose the correct IDMs. With this information, the instructional designers are better equipped to select more suited IDMs for their intended purpose and implement them correctly.

However, no IDMs that catered to the affective aspects of training were identified during the literature study. IDMs that address affective aspects of a skill can be useful in domains like therapy.

The findings of the literature study are promising and show high potential of AR to support training of complex skills. The overarching goal of the literature study is to explore how complex skills can be trained effectively and efficiently. Effective and efficient training can be achieved when students practice deliberately (Ericsson and Harwell, 2019). Thus, the IDMs were further analysed according to the Four Components Instructional Design (4CID) model which fosters deliberate practice (Neelen and Kirschner, 2016; Sarfo and Elen, 2006). The basic assumption of the 4C/ID model is that all complex learning can be represented in terms of its four components. The analysis of IDMs shows that IDMs are capable of supporting all the four components of the model and therefore, this result suggests that AR can be used to facilitate deliberate practice of complex skills with 4C/ID based approach.

7.1.2 [RQ2] How can design patterns be systematically implemented in AR to support deliberate practice?

RQ2 was addressed by shaping the ID4AR framework and the study reported in chapter 4 which utilised it to design the WEKIT.One prototype. The ID4AR framework encapsulates the 4C/ID model with the affordances of AR enabling designers to design AR environments for training complex skills with 4C/ID. Chapter 3 acts as a instruction manual/handbook providing a methodological approach to operationalise the ID4AR framework. However, the information provided in chapter 3 must be updated as new discoveries are made in future. For example, the list of IDMs from chapter 1 is not exhaustive and having more IDMs can extend the application scenarios of the framework. In addition, chapter 3 also provides the example case of the WEKIT.One to further support users of the ID4AR framework in a systematic implementation of the framework.

WEKIT.One was designed using the guidelines presented in chapter 3. It was intended to be used in all three WEKIT domains, namely Medicine, Aerospace and Astronaut training. Consequently to design a single solution, initially, task analysis was conducted with the mentors in each of the domains. After this step, the IDMs identified from all three domains were implemented in WEKIT.One. At least one IDM was allocated to each of the four components of the 4C/ID model to meet the basic assumption of the 4C/ID model. This allows 4C/ID based instructional design to be administered in all the three domains. This resulted in two core user profiles in the WEKIT.One prototype, namely the "Expert mode" and the "student mode". The "Expert mode" for mentors is used to create an expert model and the "Student mode" for students is used learn from the created expert model. Thus, the WEKIT.One prototype was tested both with mentors and students in all three domains.

Chapter 4 reports on the "Validation of IDMs" and the usability of the user study conducted in WEKIT with WEKIT.One prototype. Other parts of the work were reported by Vovk et al. (2018), who measured any potential simulation sickness that was caused by the hardware and the software in WEKIT.One. They concluded that it causes minimal sickness, under the limited use of at most one hour, and that any such inconveniences should be alleviated as AR technology improves to provide more fluid user experience. Similarly, Xue et al. (2019) found the user satisfaction to be acceptable for both mentors and students. They also indicated that gender, age, education level, and roles of students or mentors do not have any effect on user satisfaction and that satisfaction increases when users have higher computer skills.

"Validation of IDMs" was considered an important step to ensure desired outcomes in future studies of WEKIT. IDMs are abstract and do not define its implementation on the application level. For example, the IDM "Directed focus", only defines the main functionality as "Visual pointer for expert-determined relevant objects outside the visual area". How the visual pointer itself is presented/drawn to students will vary according to the designer or the domain of application. In this context, what is important is that the visual pointer is able to direct the focus of students towards the relevant area. Therefore, the study in chapter 4 was performed to test if mentors and students perceived that the implementations of IDMs in WEKIT.One satisfactorily met the intended functionalities of IDMs as defined in chapter 3.

The study was executed in three WEKIT domains in a sequential manner. Feedback from each session was incorporated before the consecutive session, thus improving the usability. This improvement was more evident in the recorder/expert view which was more sophisticated to operate with various functionalities for recording expert model. In comparison, the student view leads the student through the learning process using the expert model and therefore, the interactions are simpler. The students usability score for the player was similar and in the acceptable range in all three sessions.

A similar pattern was observed in the scores from the IDM validation. While both mentors and students found that the implementation of IDMs in WEKIT.One satisfactorily met their intended functionality, the mentors in the first session rated the IDM questionnaire items lower to their counter parts in other sessions. This finding suggests that usability may be a key factor that can influence the efficacy of the IDMs, as no changes to the implementation of IDM was made from one session to another. In addition, the perceived usefulness can vary based on the context and usefulness of the IDM for the skill being trained. The IDM "Directed focus" was rated highly in the the first session (aircraft maintenance) as compared to the others. The task in the first session required the participants to walk around the aeroplane which was a large area and thus, knowing where to go and look at next was a vital part of the task. In comparison, both the consecutive sessions had tasks where the participant was stationary. The selection of the right IDMs can have more pronounced impact on the perceived usefulness and therefore, instructional designers should always involve mentors in the design process as well. To conclude, the methodological approach provided in chapter 3 acts as guideline for the instructional designer but it does not replace key steps of the 4C/ID model. As such, it is vital to conduct extensive task analysis of the domain to select the most suitable IDMs and also implement them in a meaningful manner. Therefore, involvement of mentors is crucial throughout the whole design process.

7.1.3 [RQ3]How can expert performance be modelled and evaluated?

Chapter 5 reports on the findings of the "expert model evaluation" study conducted with the WEKIT.One prototype at the later phase of the WEKIT project. Other than the "expert model evaluation", the usability, simulator sickness and the user satisfaction were also measured which were all found to be acceptable. The objective of the "expert model evaluation" was to evaluate the expert model recorded with the WEKIT.One prototype. This study was also conducted in the three domains of WEKIT but unlike the study reported in Chapter 4, the sessions were not sequential but were conducted in parallel. In this study, a mentor from each domain created an expert model for their respective domains using the recorder/expert view of the WEKIT.One prototype. To do so, the mentor and an instructional designer involved in the design of WEKIT.One, together, broke down the learning task into concrete steps along with the required supportive and procedural information. The instructional designer in this step is optional but was used to make sure that the expert model was recorded optimally. For each of these learning steps, depending on the type of skill (cognitive, motor etc.) one/several suitable IDM/s was/were selected which the mentor could use in the WEKIT.One for recording the expert model. If the mentor considered that a certain step required repetition, he/she could also use IDMs that support the part task practice component of the 4C/ID model.

After this initial planning phase, the mentor recorded the expert model with out any assistance. Furthermore, no manual processing of the expert model was done after it was recorded and the model was loaded directly into the student view/player for training. 61 mentors, other than the ones who created the expert model, from the three domains participated in the study. These mentors used the student view to explore the expert model before responding to the expert model evaluation questionnaire. The results show positive response from the mentors in all three domains indicating the usefulness of the expert model for training. However, the mentors were not aware of the items of the questionnaire beforehand and needed to rely on their memory to answer them, which could have affected the results. There was also no significant difference between the domains which signifies that the prototype and therefore the ID4AR framework, can be used in various different domains for recording expert models with AR.

A paper-based post-test on knowledge assessment was administered among the students in the treatment and the control group. There were no significant difference between the scores of the two groups in all three domains. This result show that training with the expert model can be equally effective compared to the paper based means, despite the overhead effort required to operate the WEKIT.One. However, no pretest was conducted to factor for the previous knowledge among students which could have affected the outcome of the study. Overall, the findings indicate that the expert model recorded with WEKIT.one prototype, and by extension the ID4AR framework, is useful from perspectives of both the mentor and students.

However, most mentors and students only used the WEKIT.One prototype for one hour and there is a steep learning curve associated with operating it. Therefore, long longitudinal studies which give the participants enough time to adapt to AR interactions are needed. But AR systems are expensive and it is difficult to provide technical support to participants for long periods of time. These factors demotivates the prospect of assigning a hardware per participant for the purpose of long longitudinal studies. Additionally, requiring participants to use Hololens, which is considerably heavy, for a longer duration can cause simulation sickness (Vovk et al., 2018). As technology improves, such scenarios become more feasible to conduct.

7.1.4 [RQ4]How can feedback be designed without imposing high mental effort on students?

Results from the chapter 4 and 5 conclude that the applications designed with ID4AR are a competent tool for mentors to record expert models. Such applications can also use the expert model to support training by providing guidance and feedback using the expert model. The ID4AR framework depends on the 4C/ID model for providing general guidelines on how to provide supportive and procedural information as guidance. However, the 4C/ID model mostly relies on the human mentor to provide immediate feedback and therefore, it is unclear how an AR system must provide immediate corrective feedback. Not only can AR provide immediate feedback on the students action in a more rapid and accurate manner, it can also make use of various modalities (Schneider et al., 2017). On the other hand, improper design of such immediate feedback can increase the required mental effort, which is an undesirable effect for deliberate practice of novices (Rikers et al., 2004; Ericsson et al., 2007). To investigate the design of feedback and its consequent effect on the mental effort, the "Calligraphy trainer" was developed using the ID4AR framework. The study in chapter 6 reports on the usability study and the "Mental effort" study of the Calligraphy trainer prototype. The usability was found to be acceptable, and therefore, is unlikely to have affected the outcome of mental effort measures (Anderson, 2017), while the mental effort study investigates the effect of feedback design using the calligraphy trainer on the mental effort of calligraphy students.

The study investigates the mental effort imposed by use of three modalities (colour, saturation and auditory) for providing feedback. The mental effort was measured by triangulating three different tools, i.e. self reported mental effort (Paas et al., 2008), pupil diameter (Holmqvist et al., 2011) and dual task methodology (Brunken et al., 2003). Generally, no significant difference in the mental effort was found between the treatment and the control group using self reported mental effort and dual task methodology across all modalities Moreover, there was also no significant difference between the two groups when all feedback was provided together. This result is promising as there is no added mental effort from the feedback, however the insignificance could be the result of fewer participants in the study. In contrast, according to the pupil diameter, there was significant difference between the two groups for all modalities with the treatment group experiencing slightly higher mental effort, including the combined scenario when all feedback was provided at the same time. This is also a expected phenomena as processing a feedback usually requires additional effort. Danna and Velay (2015) suggested to use auditory modality for feedback to optimise such additional efforts but the results are inconclusive from this study. The auditory modality which, unlike the other modalities, does not share the visual channel and has the potential to optimise mental effort (Mayer, 2005). This highlights the importance of further exploration of feedback design in multimodal systems like AR.

Processing feedback takes time and effort for novices. In chapter 6, the treatment group took considerably longer than the control group to finish the task as the treatment group needed to react to the feedback provided. This is a desired effect in novices which helps the users to practice consciously. But such long duration can also be a result of non-optimised feedback. For example, use of secondary modality such as the auditory modality is recommended (Mayer and Moreno, 2003) for providing feedback but does not always lead to optimised mental effort. Complex auditory feedback such as verbal feedback demand longer attention span (Teulings and Schomaker, 1993) and can have negative effect on the mental effort. Simple non verbal feedback which requires minimal attention span can be used, effectively, by employing different strategies such as sonification (Barrass and Kramer, 1999). In addition, the findings also suggest the use of a single modality for feedback on a single parameter to simplify the processing of the feedback. In cases where more than one modality or types of feedback must be used, only one feedback should be given at a time and ample duration should be provided before giving another feedback to allow complete processing of the feedback given (Schneider et al., 2017). In conclusion, the findings of the study reported in chapter 6 signify the importance of proper design of immediate feedback and also, the need for further exploration. For example, one such step can be to investigate the types of mental effort (Brunken et al., 2003) involved in providing immediate feedback. Detailed knowledge on how the total mental effort is divided can provide more information for feedback design.

7.2 Limitations

All work presented in this thesis up to chapter 5 was done in the context of WEKIT project. This not only bound the studies to the three domains of WEKIT but also to its requirements, time and resource constraints. One of such limitation was the use of WEKIT.One prototype for all the three WEKIT domains. It may have been optimal to develop a single application for each of the domain using the framework. The constrains of the project also were applicable for study design. Factors like limited access to participants etc. may have affected the outcomes of the studies in chapter 4 and 5.

The limitations of the current technology may also have affected the findings of the studies. AR systems are far from the norm and are not yet ready for consumer use. AR devices such as Hololens are expensive, difficult to operate and not user-friendly enough. For example, the participants mobility in performing certain actions such as looking down or working in poorly lit places was limited by the Hololens. In addition to Hololens falling off from the participant's head, other issues such as frequent loss of tracking, rapid depletion of battery in complex tracking situations etc, were observed during the studies. Consequently, participants in chapter 4 and 5 could use the prototype for only one hour at most which may have affected the outcomes of the study. Moreover, there is a steep learning curve associated with operating WEKIT.One and perhaps, one hour may not have been enough to be properly exposed.

While AR is not a new medium, it is still in its infancy. The technology is rapidly advancing but sensors can only measure certain aspects of a task with certain accuracy. For example, inferring a complex cognitive processes with observable data is a challenging task. WEKIT.One does not automatically record mentor's cognitive process and instead, required mentors to think aloud during demonstration which, along with the complexity of operating the system, could have affected the quality of the expert model recorded in chapter 5. Moreover, high level data such as verbal data, currently cannot be used by machines to automatically provide feedback.

Several limitations exist from the core design of the ID4AR framework. The ID4AR framework extends the 4C/ID model for training complex skills with AR. In doing so it also inherits some of its limitations. For example, 4C/ID does not outline the optimal situation to use the framework (van Merriënboer and Kester, 2014). Complex skills are found in many domains and choosing which framework to use can depend on many other factors such as resources and available technology. Furthermore, in the context of the ADDIE model (Branch, 2009) which is a generic model for instructional design, the framework does not address critical steps such as development and evaluation.

The list of IDMs included in the design process of the framework is limited by the scope of the search of literature in chapter 1. The framework was designed to actively keep mentors involved in training and therefore, only included IDMs which relied on mentors to create expert models. Chapter 1 and 2 also do not address the effectiveness of IDMs individually or together as a system for training complex skills. While individual IDMs may or may not have been tested for their effectiveness by the original author, the combined effect of different IDMs will not always be the sum effect. Therefore, it is vital that instructional designers work closely with the mentors through out the application's life-cycle.

Limitations due to the COVID-19 restrictions also affected this thesis. Students and mentors could no longer take part in the planned study as they needed to be physically available and share hardware devices. The study was intend to explore further on the feedback design building on the finding of study reported in chapter 6 but could no longer be executed which brought the thesis to an early closure.

7.3 Implications for practice

In terms of practice, this thesis presents a novel, practical framework (ID4AR) and guidelines for instructional designers who wish to develop effective AR applications for training complex skills in various domains. It bridges the gap between the theoretical work and actual practice in training of complex skills. The ID4AR framework simplifies the use of 4C/ID model for designing AR based training application. The work in thesis also provides practical use cases which serves as an example for designers to develop and implement training applications with the ID4AR framework.

AR may soon be used in formal training as a norm. While online distant education has been around for a while, its counter part for practical training scenarios are only slowly taking roots. In the near future, the AR based systems similar to the prototypes developed in this project will facilitate such distant training, which at the moment is mostly limited to linear mediums such as videos. The work presented in this thesis serves as a stepping stone for researchers and application developers to realise this potential of AR.

Beyond education and research, such AR applications can take root in industrial applications. The prototypes in the project can facilitate industrial use cases such as assembly or repair manual. The expert model created can be uploaded and shared via cloud which the customers can download. This can eliminate the cost for printing paper manuals or developing specific applications/guidelines for each assembly task.

It can instead be quickly recorded by a human expert and shared.

7.4 Implications for future research

A natural course of future work, based on the design based approach, is to explore the application of the framework in new domains by developing new prototypes and/or extending existing ones. Such explorations will strengthen the validity of the framework, which is the key contribution of the research work presented in this thesis, and provide more design and application use cases. It can potentially also reveal limitations that we, in the duration of our research, could not foresee. More IDMs must be classified in the framework to allow more domains and their training requirements to be addressed. Exploring new domains can not only contribute to the list of IDMs but a new IDM can also be designed. Also, the results presented in chapter one shows a shortage of IDMs that support affective aspects which are needed to train complex skills. As AR technology advances, such IDMs will be more feasible and be able to augment cognitive and affective processes.

As AR is rapidly evolving, it is predicted to be adopted by the general population for daily use, much like its Virtual Reality (VR) brethren. Several works have been conducted in the WEKIT project in these regards. For example, "Technology Acceptance Model for Augmented Reality and Wearable Technologies" (TAMARA) (Guest et al., 2018) was developed to measure the sentiment of the users to accept AR in their daily lives. This model was used to evaluate WEKIT.One's user acceptance in (Guest et al., 2018) along with its user experience (Xue et al., 2019). Such works are vital for establishing AR as the media of future. Conducting more user studies in various domains with new prototypes will continuously contribute to the user experience that will eventually be accepted by the general users.

Evolving technology also brings new affordances for this powerful medium. AR as a media encompasses many technologies under its umbrella which provide a collective experience we have come to expect from AR. Recently artificial intelligence and machine learning examples have seen many successful use in training. These technologies can potentially complement AR based training of complex skills. Machine learning models can be trained to provide intelligent feedback (Di Mitri, 2019) according to the student's performance which can personalise training. This puts higher emphasis on the students progress and personalization of practise and increases the prospect of AR to support deliberate practice (Ericsson and Harwell, 2019). More affordances can be built to reduce even the workload of the mentor. Artificial intelligence can help mentors execute complex tasks such as orchestration by recommending types of practice based on the students' data.

The prototypes used in the studies reported in this thesis were built using the ID4AR framework which borrows 4C/IDs claim that it supports deliberate practice but does not add any strategies for supporting deliberate practice. AR is a powerful medium which puts the user in the centre of the interaction. This allows powerful learning strategies such as interactive story telling and games to be built around the user in authentic settings, which can further support deliberate practice(Koster, 2013). Technological affordances of AR will support meta-cognitive tools (Ibáñez et al., 2014) for more conscious practice and also help maintain motivation required to practice deliberately. In addition, future work can also focus on conducting studies to investigate ID4AR's claim of supporting deliberate practice and mastery of skills by conducting long longitudinal studies, as deliberate practice happens over repeated sessions.

The major part of the research reported in this thesis focuses on individual training as an expert or a student. However, complex skills can be collaborative in nature and training of a complex skill can also be done collaboratively (Day et al., 2007). Garzón et al. (2020) even argues that the highest impact of AR in education was seen in collaborative scenarios. AR can support both remote collaboration (Yoon et al., 2019) and co-located collaboration (Wells and Houben, 2020). AR technologies enable new methods of collaboration by creating a shared conceptual space. Sharing a conceptual space allows co-located or remote collaborators to work on the idea together in a manner that allows better communication between the participants. For example, by sharing a conceptual space, AR can provide affordances such as telepresence which allows mentors to train novices across physical barriers effectively as demonstrated by Chinthammit et al. (2014).

In conclusion, the presented research has highlighted the potential of AR to support deliberate practice of complex skills. AR hosts an ever growing collection of affordances that can give birth to new IDMs which in turn can have a significant effect on the training outcome. Importantly. the findings have shown that AR is a growing medium that is poised to enrich how we interact and learn in future, further stressing the importance vast amount of researched that is needed in various domains to realise the full potential of AR.

Acknowledgement

It was a long tiring journey that started from a small town in the foothills of Nepal to writing this thesis on the other part of the world. What started as tipsy joking threats from my father, warning to make me a pig surgeon (butcher) if i did not become a doctor (medical), has actually successfully pushed me to be here today (not the doctor you wanted but its something, please let me come home!!). At the same time, apologies to my mother who wanted me to join the military or be rich, both of which will probably not be fulfilled in this life time. I want to thank both my parents for being a yin and yang in my life, my dad for continuously persuading me to strive for more and my mum for making sure that i am always sceptical.

I am grateful to my team of promoters who guided me patiently (very...very patiently) for which i would like to offer them my sincere gratitude. Similarly, especial thanks to every one at the Open university and those who left prior to me, for making this journey enjoyable and educational. I hope to live up to your expectations of me someday and will continue to better myself. Except for my coffee school teachers, I am sorry I failed you.

Thanks to everyone at the WEKIT project and other collaborators who contributed to this research in their own ways. Finally, thanks to all the teachers who have taught me everything i know, the term teacher being used loosely in the unofficial context.

Bibliography

Ackermans, K., Rusman, E., Brand-Gruwel, S., and Specht, M. (2016). A first step towards synthesizing rubrics and video for the formative assessment of complex skills. In *International Computer Assisted Assessment Conference*, pages 1–10. Springer.

Ahmmad, S. N. Z., Su, E., Yeong, C., and Narayanan, A. L. T. (2014). Experimental study of surgeon's psychomotor skill using sensor-based measurement. *Procedia Computer Science*, 42:130–137.

Allain, K., Dado, B., Van Gelderen, M., Hokke, O., Oliveira, M., Bidarra, R., Gaubitch, N. D., Hendriks, R. C., and Kybartas, B. (2015). An audio game for training navigation skills of blind children. In *2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*, pages 1–4. IEEE.

Altimira, D., Clarke, J., Lee, G., Billinghurst, M., Bartneck, C., et al. (2017). Enhancing player engagement through game balancing in digitally augmented physical games. *International Journal of Human-Computer Studies*, 103:35–47.

Anderson, L. (2017). Cognitive load and it's impact on usability. `https://www.aytech.ca/blog/cognitive-load/`.

Araki, A., Makiyama, K., Yamanaka, H., Ueno, D., Osaka, K., Nagasaka, M., Yamada, T., and Yao, M. (2017). Comparison of the performance of experienced and novice surgeons: measurement of gripping force during laparoscopic surgery performed on pigs using forceps with pressure sensors. *Surgical Endoscopy*, 31(4):1999–2005.

Asadipour, A., Debattista, K., and Chalmers, A. (2017). Visuohaptic augmented feedback for enhancing motor skills acquisition. *The Visual Computer*, 33(4):401–411.

Asselborn, T., Gargot, T., Kidziński, Ł., Johal, W., Cohen, D., Jolly, C., and Dillenbourg, P. (2018). Automated human-level diagnosis of dysgraphia using a consumer tablet. *NPJ Digital Medicine*, 1(1):1–9.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., and MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6):34–47.

Bacca, J., Baldiris, S., Fabregat, R., Graf, S., and Kinshuk (2014). Augmented reality trends in education: A systematic review of research and applications. *Journal of Educational Technology & Society*, 17(4):133–149.

Bangor, A., Kortum, P., and Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3):114–123.

Barnes, R. (1987). Surgical handicraft: teaching and learning surgical skills. *American Journal of Surgery*, 153(5):422—427.

Barrass, S. and Kramer, G. (1999). Using sonification. *Multimedia Systems*, 7(1):23–31.

Baur, B., Fürholzer, W., Marquardt, C., and Hermsdörfer, J. (2009). Auditory grip force feedback in the treatment of writer's cramp. *Journal of Hand Therapy*, 22(2):163–171.

Benedetti, F., Volpi, N. C., Parisi, L., and Sartori, G. (2014). Attention training with an easy–to–use brain computer interface. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 236–247. Springer.

Bloom, B. and Sosniak, L. (1985). *Developing Talent in Young People*. Ballantine Books.

Bordegoni, M., Ferrise, F., Carrabba, E., Di Donato, M., Fiorentino, M., and Uva, A. E. (2014). An application based on augmented reality and mobile technology to support remote maintenance. In *Conference and Exhibition of the European Association of Virtual and Augmented Reality*, volume 1, pages 131–135.

Borges, L. R., Martins, F. R., Naves, E. L., Bastos, T. F., and Lucena, V. F. (2016). Multimodal system for training at distance in a virtual or augmented reality environment for users of electric-powered wheelchairs. *IFAC-PapersOnLine*, 49(30):156–160.

Bower, M. and Sturman, D. (2015). What are the educational affordances of wearable technologies? *Computers & Education*, 88:343–353.

Branch, R. M. (2009). *Instructional design: The ADDIE approach*, volume 722. Springer Science & Business Media.

Brooke, J. (2013). Sus: a retrospective. *Journal of Usability Studies*, 8(2):29–40.

Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability Evaluation in Industry*, 189(194):4–7.

Brunken, R., Plass, J. L., and Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, 38(1):53–61.

Buń, P., Górski, F., Wichniarek, R., Kuczko, W., and Zawadzki, P. (2015). Immersive educational simulation of medical ultrasound examination. *Procedia Computer Science*, 75:186–194.

Carey, B. (2014). How do you get to carnegie hall? talent. `https://www.nytimes.com/2014/07/15/science/which-matters-more-talent-or-practice.html`.

CBINSIGHTS (2016). AR/VR funding in 2016 already sees 85% growth on 2015. `https://www.cbinsights.com/research/augmented-virtual-reality-funding-trends-q2-2016/`.

Chang, Y.-J., Kang, Y.-S., Chang, Y.-S., and Liu, H.-H. (2015). Arcoach 2.0: Optimizing a vocational prompting system based on augmented reality for people with cognitive impairments. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pages 313–314.

Chen, X., Xu, L., Wang, Y., Wang, H., Wang, F., Zeng, X., Wang, Q., and Egger, J. (2015). Development of a surgical navigation system based on augmented reality using an optical see-through head-mounted display. *Journal of Biomedical Informatics*, 55:124–131.

Chia, F.-Y. and Saakes, D. (2014). Interactive training chopsticks to improve fine motor skills. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*, pages 1–4.

Chinthammit, W., Merritt, T., Pedersen, S., Williams, A., Visentin, D., Rowe, R., and Furness, T. (2014). Ghostman: augmented reality application for telerehabilitation and remote instruction of a novel motor skill. *BioMed Research International*, 2014.

Choi, H., Park, Y., Lee, S., Ha, H., Kim, S., Cho, H. S., and Hong, J. (2017). A portable surgical navigation device to display resection planes for bone tumor surgery. *Minimally Invasive Therapy & Allied Technologies*, 26(3):144–150.

Chong, Y. B., Bennecer, A., Hagglund, F., Siddiqi, S., Kappatos, V., Selcuk, C., and Gan, T.-H. (2015). A new synthetic training environment system based on an ict-approach for manual ultrasonic testing. *Measurement*, 71:11–22.

Cirulis, A. and Liepina, E. (2014). Designing an interactive and augmented 3d environment with passive tactile feedback for veterinary training. In *International Conference on Augmented and Virtual Reality*, pages 442–449. Springer.

Collins, A., Brown, J. S., and Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator*, 15(3):6–11.

Collins, A., Brown, J. S., and Newman, S. E. (1988). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. *Thinking: The Journal of Philosophy for Children*, 8(1):2–10.

Condino, S., Viglialoro, R. M., Fani, S., Bianchi, M., Morelli, L., Ferrari, M., Bicchi, A., and Ferrari, V. (2016). Tactile augmented reality for arteries palpation in open surgery training. In *International Conference on Medical Imaging and Augmented Reality*, pages 186–197. Springer.

Dalle Mura, M., Dini, G., and Failli, F. (2016). An integrated environment based on augmented reality and sensing device for manual assembly workstations. *Procedia Cirp*, 41:340–345.

Danna, J. and Velay, J.-L. (2015). Basic and supplementary sensory feedback in handwriting. *Frontiers in Psychology*, 6:169.

Daponte, P., De Vito, L., Riccio, M., and Sementa, C. (2014). Design and validation of a motion-tracking system for rom measurements in home rehabilitation. *Measurement*, 55:82–96.

Datcu, D., Cidota, M., Lukosch, S., Oliveira, D. M., and Wolff, M. (2014). Virtual co-location to support remote assistance for inflight maintenance in ground training for space missions. In *Proceedings of the 15th International Conference on Computer Systems and Technologies*, pages 134–141.

Day, E. A., Boatman, P. R., Kowollik, V., Espejo, J., McEntire, L. E., and Sherwin, R. E. (2007). Collaborative training with a more experienced partner: Remediating low pretraining self-efficacy in complex skill acquisition. *Human Factors*, 49(6):1132–1148.

De Corte, E., for Research on Learning, E. A., Instruction, Verschaffel, L., Van Merrienboer, J., and Entwistle, N. (2003). *Powerful Learning Environments: Unravelling Basic Components and Dimensions.* Advances in learning and instruction series. Pergamon.

De Paolis, L. T., Ricciardi, F., and Giuliani, F. (2014). Development of a serious game for laparoscopic suture training. In *International Conference on Augmented and Virtual Reality*, pages 90–102. Springer.

De Ravé, E. G., Jiménez-Hornero, F. J., Ariza-Villaverde, A. B., and Taguas-Ruiz, J. (2016). Diedricar: a mobile augmented reality system designed for the ubiquitous descriptive geometry learning. *Multimedia Tools and Applications*, 75(16):9641–9663.

Di Mitri, D. (2019). Detecting medical simulation errors with machine learning and multimodal data. In *17th Conference on Artificial Intelligence in Medicine*.

Di Mitri, D., Schneider, J., Klemke, R., Specht, M., and Drachsler, H. (2019). Read between the lines: an annotation tool for multimodal data for learning. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 51–60.

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

DiSessa, A. A. and Cobb, P. (2004). Ontological innovation and the role of theory in design experiments. *The Journal of the Learning Sciences*, 13(1):77–103.

Djajadiningrat, T., Lui, P., Chao, P.-Y., and Richard, C. (2016). Virtual trainer: A low cost ar simulation of a sudden cardiac arrest emergency. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pages 607–618.

Drljević, N., Wong, L. H., and Botički, I. (2017). Where does my augmented reality learning experience (arle) belong? a student and teacher perspective to positioning arles. *IEEE Transactions on Learning Technologies*, 10(4):419–435.

Ericsson, K. A. and Harwell, K. W. (2019). Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: Why the original definition matters and recommendations for future research. *Frontiers in Psychology*, 10:2396.

Ericsson, K. A., Hoffman, R. R., and Kozbelt, A. (2018). *The Cambridge handbook of expertise and expert performance*. Cambridge University Press.

Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3):363.

Ericsson, K. A., Prietula, M. J., and Cokely, E. T. (2007). The making of an expert. *Harvard Business Review*, 85(7/8):114.

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

Fominykh, M., Wild, F., Smith, C., Alvarez, V., and Morozov, M. (2015). An overview of capturing live experience with virtual and augmented reality. In *Intelligent Environments (Workshops)*, pages 298–305.

Frenoy, R., Thouvenin, I., Soullard, Y., and Gapenne, O. (2016). CalliSmart: an Adaptive Informed Environment for Intelligent Calligraphy Training. In *The Ninth International Conference on Advances in Computer-Human Interactions (ACHI 2016)*, pages 132–137, Venice, Italy.

Frerejean, J., van Merriënboer, J. J., Kirschner, P. A., Roex, A., Aertgeerts, B., and Marcellis, M. (2019). Designing instruction for complex learning: 4C/ID in higher education. *European Journal of Education*, 54(4):513–524.

Freschi, C., Parrini, S., Dinelli, N., Ferrari, M., and Ferrari, V. (2015). Hybrid simulation using mixed reality for interventional ultrasound imaging training. *International Journal of Computer Assisted Radiology and Surgery*, 10(7):1109–1115.

Funk, M., Heusler, J., Akcay, E., Weiland, K., and Schmidt, A. (2016). Haptic, auditory, or visual? towards optimal error feedback at manual assembly workplaces. In *Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, pages 1–6.

Gallegos-Nieto, E., Medellín-Castillo, H. I., González-Badillo, G., Lim, T., and Ritchie, J. (2017). The analysis and evaluation of the influence of haptic-enabled virtual assembly training on real assembly performance. *The International Journal of Advanced Manufacturing Technology*, 89(1-4):581–598.

Garbi, I. (2020). Futureskills : Forecast skill trends, at the country level, to help governments and people prepare for a transforming world of work. - view idea. `https://ideas.unite.un.org/futureskills/Page/ViewIdea?ideaid=4284`.

Garzón, J., Baldiris, S., Gutiérrez, J., Pavón, J., et al. (2020). How do pedagogical approaches affect the impact of augmented reality on education? a meta-analysis and research synthesis. *Educational Research Review*, page 100334.

Gladwell, M. (2008). *Outliers: The story of success*. Little, Brown.

Gordienko, Y., Stirenko, S., Alienin, O., Skala, K., Sojat, Z., Rojbi, A., López Benito, J. R., Artetxe González, E., Lushchyk, U., Sajn, L., Llorente Coto, A., and Jervan, G. (2017). Augmented coaching ecosystem for non-obtrusive adaptive personalized elderly care on the basis of cloud-fog-dew computing paradigm. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 359–364.

Guest, W., Wild, F., Vovk, A., Fominykh, M., Limbu, B., Klemke, R., Sharma, P., Karjalainen, J., Smith, C., Rasool, J., et al. (2017). Affordances for capturing and reenacting expert performance with wearables. In *European Conference on Technology Enhanced Learning*, pages 403–409. Springer.

Guest, W., Wild, F., Vovk, A., Lefrere, P., Klemke, R., Fominykh, M., and Kuula, T. (2018). A technology acceptance model for augmented reality and wearable technologies. *Journal of Universal Computer Science*, 24(2):192–219.

Hahn, J., Ludwig, B., and Wolff, C. (2015). Augmented reality-based training of the pcb assembly process. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, pages 395–399.

Han, H., Han, S. H., Lim, D. H., and Yoon, S. W. (2015). Active learning, deliberate practice, and educational technology in professional education: Practices and implications. In *Handbook of Research on Educational Technology Integration and Active Learning*, pages 177–201. IGI Global.

Hattie, J. (2017). Backup of hattie's ranking list of 256 influences and effect sizes related to student achievement. Available at `https://visible-learning.org/backup-hattie-ranking-256-effects-2017/` (31/05/2020).

Haug, T., Rozenblit, J. W., and Buchenrieder, K. (2014). Movement analysis in laparoscopic surgery training. In *Proceedings of the 2014 Summer Simulation Multiconference*, pages 1–8.

Hinds, P. J. (1999). The curse of expertise: The effects of expertise and debiasing methods on prediction of novice performance. *Journal of Experimental Psychology: applied*, 5(2):205.

Hinds, P. J., Patterson, M., and Pfeffer, J. (2001). Bothered by abstraction: The effect of expertise on knowledge transfer and subsequent novice performance. *Journal of Applied Psychology*, 86(6):1232.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Horeman, T., Blikkendaal, M. D., Feng, D., van Dijke, A., Jansen, F., Dankelman, J., and van den Dobbelsteen, J. J. (2014). Visual force feedback improves knot-tying security. *Journal of Surgical Education*, 71(1):133–141.

Ibáñez, M.-B., Di-Serio, Á., Villarán-Molina, D., and Delgado-Kloos, C. (2014). Augmented reality-based simulators as discovery learning tools: An empirical study. *IEEE Transactions on Education*, 58(3):208–213.

Islam, G., Kahol, K., Li, B., Smith, M., and Patel, V. L. (2016). Affordable, web-based surgical skill training and evaluation tool. *Journal of Biomedical Informatics*, 59:102–114.

Jang, S.-A., Kim, H.-i., Woo, W., and Wakefield, G. (2014). Airsculpt: A wearable augmented reality 3d sculpting system. In *International Conference on Distributed, Ambient, and Pervasive Interactions*, pages 130–141. Springer.

Jarman, C. (1979). *The Development of Handwriting Skills: A Book of Resources for Teachers*. Primary education. Primary Education.

Jarodzka, H., Gruber, H., and Holmqvist, K. (2017). Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research*.

Jarodzka, H., Van Gog, T., Dorr, M., Scheiter, K., and Gerjets, P. (2013). Learning to see: Guiding students' attention via a model's eye movements fosters learning. *Learning and Instruction*, 25:62–70.

Juanes, J. A., Gómez, J. J., Peguero, P. D., and Ruisoto, P. (2015). Practical applications of movement control technology in the acquisition of clinical skills. In *Proceedings of the 3rd International Conference on Technological Ecosystems for Enhancing Multiculturality*, pages 13–17.

Juanes, J. A., Hernández, D., Ruisoto, P., García, E., Villarrubia, G., and Prats, A. (2014). Augmented reality techniques, using mobile devices, for learning human anatomy. In *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality*, pages 7–11.

Jucks, R., Schulte-Löbbert, P., and Bromme, R. (2007). Supporting experts' written knowledge communication through reflective prompts on the use of specialist concepts. *Journal of Psychology*, 215(4):237–247.

Kamphuis, C., Barsom, E., Schijven, M., and Christoph, N. (2014). Augmented reality in medical education? *Perspectives on Medical Education*, 3(4):300–311.

Ke, F., Lee, S., and Xu, X. (2016). Teaching training in a mixed-reality integrated learning environment. *Computers in Human Behavior*, 62:212–220.

Kersten-Oertel, M., Gerard, I. J., Drouin, S., Petrecca, K., Hall, J. A., and Collins, D. L. (2016). Towards augmented reality guided craniotomy planning in tumour resections. In *International Conference on Medical Imaging and Augmented Reality*, pages 163–174. Springer.

Khan, M. (2015). Transmitting al ardha: Traditional arab sword dance. *International Journal of Heritage in the Digital Era*, 4(1):71–85.

Kim, S. and Dey, A. K. (2016). Augmenting human senses to improve the user experience in cars: applying augmented reality and haptics approaches to reduce cognitive distances. *Multimedia Tools and Applications*, 75(16):9587–9607.

Koreeda, Y., Kobayashi, Y., Ieiri, S., Nishio, Y., Kawamura, K., Obata, S., Souzaki, R., Hashizume, M., and Fujie, M. G. (2016). Virtually transparent surgical instruments in endoscopic surgery with augmentation of obscured regions. *International Journal of Computer Assisted Radiology and Surgery*, 11(10):1927–1936.

Koster, R. (2013). *Theory of fun for game design*. O'Reilly Media, Inc.

Kowalewski, K.-F., Hendrie, J. D., Schmidt, M. W., Garrow, C. R., Bruckner, T., Proctor, T., Paul, S., Adigüzel, D., Bodenstedt, S., Erben, A., et al. (2017). Development and validation of a sensor-and expert model-based training system for laparoscopic surgery: the isurgeon. *Surgical Endoscopy*, 31(5):2155–2165.

Kritopoulou, P., Manitsaris, S., and Moutarde, F. (2016). Towards the design of augmented feedforward and feedback for sensorimotor learning of motor skills. In *Proceedings of the 3rd International Symposium on Movement and Computing*, pages 1–4.

Kwon, Y., Lee, S., Jeong, J., and Kim, W. (2014). Heartisense: A novel approach to enable effective basic life support training without an instructor. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '14, page 431–434, New York, NY, USA. Association for Computing Machinery.

Lahanas, V., Loukas, C., Smailis, N., and Georgiou, E. (2015). A novel augmented reality simulator for skills assessment in minimal invasive surgery. *Surgical Endoscopy*, 29(8):2224–2234.

Langlotz, T., Grubert, J., and Grasset, R. (2013). Augmented reality browsers: essential products or only gadgets? *Communications of the ACM*, 56(11):34–36.

Lee, Sun, S.-H., Somasundaram, S., Hu, E. S., and Lim, J. J. (2018). Composing complex skills by learning transition policies. In *International Conference on Learning Representations*.

Lee, I.-J., Chen, C.-H., and Chang, K.-P. (2016). Augmented reality technology combined with three-dimensional holography to train the mental rotation ability of older adults. *Computers in Human Behavior*, 65:488–500.

Li, H., Lu, M., Chan, G., and Skitmore, M. (2015). Proactive training system for safe and efficient precast installation. *Automation in Construction*, 49:163–174.

Limbu, B., Fominykh, M., Klemke, R., and Specht, M. (2019a). *A Conceptual Framework for Supporting Expertise Development with Augmented Reality and Wearable Sensors*, pages 213–228. Springer International Publishing, Cham.

Limbu, B., Fominykh, M., Klemke, R., Specht, M., and Wild, F. (2018a). *Supporting Training of Expertise with Wearable Technologies: The WEKIT Reference Framework*, pages 157–175. Springer Singapore, Singapore.

Limbu, B., Vovk, A., Jarodzka, H., Klemke, R., Wild, F., and Specht, M. (2019b). Wekit.one: A sensor-based augmented reality system for experience capture and re-enactment. In Scheffel, M., Broisin, J., Pammer-Schindler, V., Ioannou, A., and Schneider, J., editors, *Transforming Learning with Meaningful Technologies*, pages 158–171, Cham. Springer International Publishing.

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

Limbu, B. H., Jarodzka, H., Klemke, R., and Specht, M. (2019c). Can you ink while you blink? Assessing mental effort in a sensor-based calligraphy trainer. *Sensors (Basel, Switzerland)*, 19(14).

Limbu, B. H., Jarodzka, H., Klemke, R., Wild, F., and Specht, M. (2018c). From AR to expertise: A user study of an augmented reality training to support expertise development. *Journal of Universal Computer Science*, 24(2):108–128. `http://www.jucs.org/jucs_24_2/from_ar_to_expertise`.

Lin, C.-Y., Chen, C.-J., Liu, Y.-H., Chai, H.-C., Lin, C.-W., Huang, Y.-M., Chen, C.-W., and Lin, C.-C. (2015). Integrating motion-capture augmented reality technology as an interactive program for children. In *International Conference on Universal Access in Human-Computer Interaction*, pages 149–156. Springer.

Liu, J., Mei, J., Zhang, X., Lu, X., and Huang, J. (2017). Augmented reality-based training system for hand rehabilitation. *Multimedia Tools and Applications*, 76(13):14847–14867.

Lok, B., Chuah, J. H., Robb, A., Cordar, A., Lampotang, S., Wendling, A., and White, C. (2014). Mixed-reality humans for team training. *IEEE Computer Graphics and Applications*, 34(3):72–75.

Loup-Escande, E., Frenoy, R., Poplimont, G., Thouvenin, I., Gapenne, O., and Megalakaki, O. (2017). Contributions of mixed reality in a calligraphy learning task: Effects of supplementary visual feedback and expertise on cognitive load, user experience and gestural performance. *Computers in Human Behavior*, 75:42–49.

ManpowerGroup (2020). Closing the Skills Gap: What Workers Want. `https://go.manpowergroup.com/talent-shortage`.

Manuri, F., Sanna, A., Lamberti, F., Paravati, G., and Pezzolla, P. (2014). A workflow analysis for implementing ar-based maintenance procedures. In *International Conference on Augmented and Virtual Reality*, pages 185–200. Springer.

Matassa, A. and Morreale, F. (2016). Supporting singers with tangible and visual feedback. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 328–329.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge Handbook of Multimedia Learning*, 41:31–48.

Mayer, R. E. and Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, 38(1):43–52.

McDaniel, M. A., Schmidt, F. L., and Hunter, J. E. (1988). Job experience correlates of job performance. *Journal of Applied Psychology*, 73(2):327.

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., and Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003–2009. *Medical Education*, 44(1):50–63.

Meleiro, P., Rodrigues, R., Jacob, J., and Marques, T. (2014). Natural user interfaces in the motor development of disabled children. *Procedia Technology*, 13:66 – 75. SLACTIONS 2013: Research conference on virtual worlds – Learning with simulations.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development*, 50(3):43–59.

Merriënboer, J. J. G. v. and Paas, F. (2003). Powerful learning and the many faces of instructional design: Toward a framework for the design of powerful learning environments. In *Powerful Learning Environments: Unravelling Basic Components and Dimensions.*, Advances in learning and instruction series., pages 3–20. Pergamon/Elsevier Science Ltd, Oxford, England.

Milazzo, N., Farrow, D., and Fournier, J. F. (2016). Effect of implicit perceptual-motor training on decision-making skills and underpinning gaze behavior in combat athletes. *Perceptual and Motor Skills*, 123(1):300–323.

Morikawa, A., Tsuda, N., Nomura, Y., and Kato, N. (2018). Double pressure presentation for calligraphy self-training. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 199–200.

Neelen, M. and Kirschner, P. (2016). Deliberate practice: What it is and what it isn't. `https://3starlearningexperiences.wordpress.com/2016/06/21/370/`.

Neelen, M. and Kirschner, P. A. (2018). Complex skills in the workplace and what it means to design for them. `https://3starlearningexperiences.wordpress.com/2018/03/20/complex-skills-in-the-workplace-and-what-it-means-to-design-for-them/`.

Olwal, A., Gustafsson, J., and Lindfors, C. (2008). Spatial augmented reality on industrial cnc-machines. In *The Engineering Reality of Virtual Reality 2008*, volume 6804, page 680409. International Society for Optics and Photonics.

Onishi, K., Mizushino, K., Noborio, H., and Koeda, M. (2014). Haptic ar dental simulator using z-buffer for object deformation. In *International Conference on Universal Access in Human-Computer Interaction*, pages 342–348. Springer.

Oyekan, J., Prabhu, V., Tiwari, A., Baskaran, V., Burgess, M., and Mcnally, R. (2017). Remote real-time collaboration through synchronous exchange of digitised human–workpiece interactions. *Future Generation Computer Systems*, 67:83–93.

Paas, F., Ayres, P., and Pachman, M. (2008). Assessment of cognitive load in multimedia learning. *Recent Innovations in Educational Technology That Facilitate Student Learning, Information Age Publishing Inc., Charlotte, NC*, pages 11–35.

Papagiannis, H. (2017). *Augmented human: How technology is shaping the new reality.* O'Reilly Media, Inc.

Park, K., Kihl, T., Park, S., Kim, M.-J., and Chang, J. (2016). Narratives and sensor driven cognitive behavior training game platform. In *2016 IEEE 14th International Conference on Software Engineering Research, Management and Applications (SERA)*, pages 125–131. IEEE.

Patterson, R. E., Pierce, B. J., Bell, H. H., and Klein, G. (2010). Implicit learning, tacit knowledge, expertise development, and naturalistic decision making. *Journal of Cognitive Engineering and Decision Making*, 4(4):289–303.

Perlini, S., Salinaro, F., Santalucia, P., and Musca, F. (2014). Simulation-guided cardiac auscultation improves medical students' clinical skills: the pavia pilot experience. *Internal and Emergency Medicine*, 9(2):165–172.

Prabhu, V. A., Elkington, M., Crowley, D., Tiwari, A., and Ward, C. (2017). Digitisation of manual composite layup task knowledge using gaming technology. *Composites Part B: Engineering*, 112:314–326.

Radu, I., Doherty, E., DiQuollo, K., McCarthy, B., and Tiu, M. (2015). Cyberchase shape quest: pushing geometry education boundaries with augmented reality. In *Proceedings of the 14th International Conference on Interaction Design and Children*, pages 430–433.

Rane, P., Kim, H., Marcano, J. L., and Gabbard, J. L. (2016). Virtual road signs: Augmented reality driving aid for novice drivers. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 60, pages 1750–1754. SAGE Publications Sage CA: Los Angeles, CA.

Rikers, R. M., Van Gerven, P. W., and Schmidt, H. G. (2004). Cognitive load theory as a tool for expertise development. *Instructional Science*, 32(1-2):173–182.

Roads, B., Mozer, M. C., and Busey, T. A. (2016). Using highlighting to train attentional expertise. *PloS one*, 11(1).

Rozenblit, J. W., Feng, C., Riojas, M., Napalkova, L., Hamilton, A. J., Hong, M., Berthet-Rayne, P., Czapiewski, P., Hwang, G., Nikodem, J., et al. (2014). The computer assisted surgical trainer: design, models, and implementation. In *Proceedings of the 2014 Summer Simulation Multiconference*, pages 1–10.

Sand, O., Büttner, S., Paelke, V., and Röcker, C. (2016). smart. assembly–projection-based augmented reality for supporting assembly workers. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 643–652. Springer.

Sanfilippo, F. (2017). A multi-sensor fusion framework for improving situational awareness in demanding maritime training. *Reliability Engineering & System Safety*, 161:12–24.

Sano, Y., Sato, K., Shiraishi, R., and Otsuki, M. (2016). Sports support system: Augmented ball game for filling gap between player skill levels. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*, pages 361–366.

Sarfo, F. K. and Elen, J. (2006). Technical expertise development in secondary technical schools: Effects of ictenhanced 4c/id learning environments. In *Fourth IEEE International Workshop on Technology for Education in Developing Countries (TEDC'06)*, pages 62–65. IEEE.

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

Schneider, J., Börner, D., Van Rosmalen, P., and 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., and Drachsler, H. (2018). Multimodal learning hub: A tool for capturing customizable multimodal learning experiences. In *European Conference on Technology Enhanced Learning*, pages 45–58. Springer.

Schraffenberger, H. and van der Heide, E. (2016). Multimodal augmented reality: the norm rather than the exception. In *Proceedings of the 2016 workshop on Multimodal Virtual and Augmented Reality*, pages 1–6.

Sebillo, M., Tortora, G., Vitiello, G., Paolino, L., and Ginige, A. (2015). The use of augmented reality interfaces for on-site crisis preparedness. In *International Conference on Learning and Collaboration Technologies*, pages 136–147. Springer.

See, Z. S., Billinghurst, M., Rengganaten, V., and Soo, S. (2016). Medical learning murmurs simulation with mobile audible augmented reality. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications*, SA '16, New York, NY, USA. Association for Computing Machinery.

Shekhar, R., Dandekar, O., Bhat, V., Philip, M., Lei, P., Godinez, C., Sutton, E., George, I., Kavic, S., Mezrich, R., et al. (2010). Live augmented reality: a new visualization method for laparoscopic surgery using continuous volumetric computed tomography. *Surgical Endoscopy*, 24(8):1976–1985.

Sigrist, R., Rauter, G., Riener, R., and Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic Bulletin & Review*, 20(1):21–53.

Sousa, L., Alves, R., and Rodrigues, J. (2016). Augmented reality system to assist inexperienced pool players. *Computational Visual Media*, 2(2):183–193.

Specht, M., Limbu, B. H., and Schneider, J. (2019). *Sensors for Seamless Learning*, pages 141–152. Springer Singapore, Singapore.

Stunt, J., Kerkhoffs, G., Horeman, T., van Dijk, C., and Tuijthof, G. (2016). Validation of the passport v2 training environment for arthroscopic skills. *Knee Surgery, Sports Traumatology, Arthroscopy*, 24(6):2038–2045.

Such, M., Ward, C., Hutabarat, W., and Tiwari, A. (2014). Intelligent composite layup by the application of low cost tracking and projection technologies. *Procedia CIRP*, 25:122 – 131. 8th International Conference on Digital Enterprise Technology - DET 2014 Disruptive Innovation in Manufacturing Engineering towards the 4th Industrial Revolution.

Sun, X., Byrns, S., Cheng, I., Zheng, B., and Basu, A. (2017). Smart sensor-based motion detection system for hand movement training in open surgery. *Journal of Medical Systems*, 41(2):24.

Szulewski, A., Roth, N., and Howes, D. (2015). The use of task-evoked pupillary response as an objective measure of cognitive load in novices and trained physicians: a new tool for the assessment of expertise. *Academic Medicine*, 90(7):981–987.

Teulings, H.-L. and Schomaker, L. R. (1993). Invariant properties between stroke features in handwriting. *Acta Psychologica*, 82(1-3):69–88.

Thorpe, M. (2013). *Modern Calligraphy: Everything You Need to Know to Get Started in Script Calligraphy.* St. Martin's Publishing Group.

Tokuyasu, T., Okamura, W., Kusano, T., Inomata, M., Shiraishi, N., and Kitanou, S. (2014). Training system for endoscopic surgery by using augmented reality and forceps control devices. In *2014 Ninth International Conference on Broadband and Wireless Computing, Communication and Applications*, pages 541–544. IEEE.

Tong, Y., Wang, Y., Chen, J., and Chen, C. (2016). A small scene assistant maintenance system based on optical see-through augmented reality. In *Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry-Volume 1*, pages 155–158.

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

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

Van Merriënboer, J. J. and Kirschner, P. A. (2017). *Ten steps to complex learning: A systematic approach to four-component instructional design.* Routledge.

van Merriënboer, J. J. G. and Kester, L. (2014). *The Four-Component Instructional Design Model: Multimedia Principles in Environments for Complex Learning*, page 104–148. Cambridge Handbooks in Psychology. Cambridge University Press, 2 edition.

Vovk, A., Wild, F., Guest, W., and Kuula, T. (2018). Simulator sickness in augmented reality training using the microsoft hololens. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–9.

Wang, X., Ong, S., and Nee, A. Y.-C. (2016). Multi-modal augmented-reality assembly guidance based on bare-hand interface. *Advanced Engineering Informatics*, 30(3):406–421.

Wei, Y., Yan, H., Bie, R., Wang, S., and Sun, L. (2014). Performance monitoring and evaluation in dance teaching with mobile sensing technology. *Personal and Ubiquitous Computing*, 18(8):1929–1939.

Wells, T. and Houben, S. (2020). Collabar–investigating the mediating role of mobile ar interfaces on co-located group collaboration. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Western, C., Hristov, D., and Schlosser, J. (2015). Ultrasound imaging in radiation therapy: From interfractional to intrafractional guidance. *Cureus*, 7(6).

Wu, H.-K., Lee, S. W.-Y., Chang, H.-Y., and Liang, J.-C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education*, 62:41–49.

Xue, H., Sharma, P., and Wild, F. (2019). User satisfaction in augmented reality-based training using microsoft hololens. *Computers*, 8(1):9.

Yoon, B., Kim, H.-i., Lee, G. A., Billinqhurst, M., and Woo, W. (2019). The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 547–556. IEEE.

Zhao, Y., Salunke, S., Leavitt, A., Curtin, K., Huynh, N., and Zeagler, C. (2016). E-archery: prototype wearable for analyzing archery release. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 908–913.

Zhu, Z., Branzoi, V., Wolverton, M., Murray, G., Vitovitch, N., Yarnall, L., Acharya, G., Samarasekera, S., and Kumar, R. (2014). Ar-mentor: Augmented reality based mentoring system. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 17–22. IEEE.

Appendices

Table 1: List of analysed studies and their implemented transfer mechanism.

Studies	Description	Sensor Used	Experts Involvement	Instructional Methods
(De Paolis et al., 2014)	Serious Game for assessment of skill in laparoscopic suture training by physical modelling of virtual environments. Gamification of surgical skill performed on virtual models using the haptic sensors for physical haptic feedback	Haptic sensors		Interactive virtual object
(Liu et al., 2017)	AR system for hand movement rehabilitation. Patients were asked to imitate a trajectory or move virtual objects	Optical marker trackers		Augmented paths, Interactive virtual objects
(Stunt et al., 2016)	Prototype passport v2 was used to train minimally invasive surgery by providing natural sensory feedback.Participants were also asked to perform a guided navigation motion	Depth camera and haptic band		Augmented paths, Haptic feedback
(Sun et al., 2017)	Motion detection system to assess hand dexterity for surgical knot tying	Leap motion	Recorded expert hand movement path to use as comparison for assessment	Augmented paths, Feedback
(Hahn et al., 2015)	Support and training for assembly of printed board circuit by highlighting component retrieval location	Smart glass and QR code tracking camera		Directed focus, Highlight object of interest
(Haug et al., 2014)	Analysis of laparoscopic surgical skills by capturing motion and translation in 3D space with the aim of classifying calculated features of the performance	Wireless inertial sensor and surgical tools with inertial tracker	Recorded the expert hand movement and the movement of the surgical tools including their translation	Augmented path
(Rozenblit et al., 2014)	Prototype for providing cognitive and haptic feedback for guided surgical training using computer-modelled expert attributes	Embedded sensors in the surgical instruments	Models the judgement criteria used by experienced surgeons and provides scalability characteristics in three key aspects: integration of new expert opinions; integration of new evaluation metrics; and integration of new performance data, for the constant improvement of an objective scoring system	Augmented path, Haptic feedback
(Tokuyasu et al., 2014)	Training of endoscopic surgery skills.The apprentice was asked to use a surgical tool to track a Japanese letter displayed through smart glass	Optical tracker		Augmented path
(De Ravé et al., 2016)	Prototype for teaching descriptive geometry learning with spatial 3D visualisations of the complex model in a series of step-by-step exercises	Augmented reality display		3D models and animation
(Chia and Saakes, 2014)	Training fine motor skills and hand-eye coordination by imitating the expert's chopstick movements overlaid using smart glasses	Potentiometer for force and accelerometer formation	Recorded expert hand movements along with chopsticks to be overlaid	Augmented mirror
(Meleiro et al., 2014)	Assisting children with motor impairments for walking with a therapy for a sequence of poses	Kinect for posture	Interview to understand the therapy	Augmented mirror
(Khan, 2015)	prototype for capturing expert performing traditional Arab dance to create a teaching system	Kinect	Recording expert demonstration	Augmented mirror

Continuation of Table 1

Studies	Description	Sensor Used	Experts Direct Involvement	Instructional Methods
(Daponte et al., 2014)	Prototype for calculation of range of motion for elbow joint for rehabilitation programme which was stored and forwarded to the expert	Kinect	Support system for the expert to design therapies	Augmented mirror
(Lin et al., 2015)	Prototype to investigate the effects of game elements intervention on the jumping performance of children	Kinect		Augmented mirror
(Wei et al., 2014)	Designed to evaluate and monitor dance performance using the expert performance as a benchmark	Use of accelerometer for motion and rhythm	Captured the expert performance	Augmented mirror
(Jang et al., 2014)	Training platform for sculpting on a virtual model with physical hand and finger movements	Depth sensing camera	Interview of the experts to determine the performance criteria	Interactive virtual object
(Ke et al., 2016)	Prototype for teaching training and classroom management in a mixed reality environment with non-player characters designed to intervene as students	Kinect, Eye tracker		Directed focus, Interactive virtual objects
(Radu et al., 2015)	Teaching geometry with a game that required children to build a bridge across the river for the game character to cross. Physical blocks were also provided to support spatial	Tablet camera		Interactive virtual objects, 3D models and animation
(Juanes et al., 2014)	Simple prototype that allowed the learner to practise laparoscopic surgery by using gestures to interact with the virtual object	Leap motion		Interactive virtual object
(Kritopoulou et al., 2016)	Prototype for teaching pottery-making using augmented feed forward and feedback to train sensorimotor skills	Depth camera, leap motion, Projector	Expert operational model which encapsulated all the information of the gesture obtained from the expert. Therefore, the model includes the subtle details of the quality of the gesture, which can distinguish an Expert from a Learner	Augmented path, Feedback
(Lok et al., 2014)	Team training simulation with non-player characters for training collaboration in medical situations	Kinect and microphone		Interactive virtual objects
(Roads et al., 2016)	Prototype to train fingerprint-matching skill by high lighting relevant areas and using fixation to assess the performance	Eye tracking	Use of expert model	Highlight objects of Interest
(Milazzo et al., 2016)	Participants were semi-professional players who were required to watch the video of a player making decisions in game			Point of view video
(Djajadiningrat et al., 2016)	Low-cost Augmented Reality simulator for sudden cardiac arrest emergency training by following a sequence of instructions in video format and the contextual information provided in the proximity of a physical mannequin	Optical tracker	Expert interviewed during design of instructions	Contextual information
(Gallegos-Nieto et al., 2017)	Designed for the evaluation of the influence of haptic-enabled virtual assembly training. Learners could use a haptic device to assemble virtual components	Haptic motors		Haptic feedback, Interactive virtual objects
(Datcu et al., 2014)	Designed for collaboration for in-flight maintenance allowing expert to assist from a different location. The expert was provided with the point of view video of the trainee with contextual icons such as arrows that were displayed to the trainee	Head mounted camera	Telepresence	Annotations Cues and clues

Continuation of Table 1

Studies	Description	Sensor Used	Experts Direct Involvement	Instructional Methods
(Lahanas et al., 2015)	Prototype for assessment of minimally invasive surgery skills using a box trainer equipped with a camera and laparoscopic tools with built-in motion sensors that allowed the learner to manipulate objects in a virtual environment	Camera,accelerometer built into the surgical tools		Interactive virtual objects
(Chen et al., 2015)	Surgical navigation system for sacroiliac joint screw insertion on a mannequin which synchronised with the 3D model being manipulated by the trainee during practice on the mannequin	Motion tracker		Interactive virtual objects
(Wang et al., 2016)	Multimodal assembly guidance system that provided procedural information and haptic feedback on error detection	Haptic band, Leap motion		Object enrichment, Contextual information, Haptic feedback
(Benedetti et al., 2014)	Prototype for attention training on adult suffering from a frontal syndrome using a brain computer interface that moved the 3D object with brain waves	EEG		Interactive virtual objects
(Onishi et al., 2014)	Dental surgery training prototype that allowed learners to practise on a virtual 3D model of teeth. The prototype also provided haptic feedback	Magnetic motion sensor, Haptic device, camera		Interactive virtual objects
(Tong et al., 2016)	Hovering hand pointer assisted the trainee by pointing at the area that needed to be addressed during the assembly of the motherboard	Camera designed to recognise the motherboard		Highlight, object of interest
(Sand et al., 2016)	Assembly assistance of Lego withprojection-based AR which displayedinstructions and highlighted the boxcontaining the next piece. Foot pedalswere used to shift to the next step	Camera, projector		Contextual information, Highlight object of interest
(Rane et al., 2016)	Prototype for training drivers' situational awareness and perception by providing driving aids on location and contextual information			Contextual information, Highlight object of interest
(Zhu et al., 2014)	AR mentoring system for maintenance	Head mounted camera, Microphone	Expert model that tracks its orientation to the object being repaired to provide feedback on the correct placement	Contextual information, Directed focus, Object enrichment, Feedback
(Chinthammit et al., 2014)	Motor rehabilitation system using visual augmentation to allow the therapist and the patient to inhabit each other's viewpoint	Head mounted camera	Telepresence of expert. The patient had to imitate the therapist's hand which was overlaid on the patient's vision	Augmented path, Point of view video
(Bordegoni et al., 2014)	Prototype to support collaboration between an expert and a novice in a maintenance task by sharing the novice's point of view video and allowing the expert to interact with the video in the form of annotations which were displayed on the novice's glasses	Head mounted camera	Telepresence. Possibility for the expert to put annotations in the novice's field of view	Annotations Cues and clues, Point of view
(Li et al., 2015)	Training system for safe precast installation in construction by improving collaboration	Motion tracking with optical camera	Telepresence. The novice's position was tracked and displayed on the screen where the expert could annotate areas of interest for the novice	Annotations, Contextual information

Continuation of Table 1

Studies	Description	Sensor Used	Experts Direct Involvement	Instructional Methods
(Sanfilippo, 2017)	Designed to train situational awareness in Maritime training by providing live guidance by the expert who provided guidance to the areas of interest	Eye tracker, Oculus rift	Experts shared the simulation view and directed the trainee to act accordingly	Point of view video, Audio instructions
(See et al., 2016)	Enhanced heart murmurs to train residents in recognising them	Mobile phone and electronic stethoscope		Object enrichment
(Manuri et al., 2014)	AR-based maintenance support system that provided the procedural information and 3D models of the objects			3D models and animation, Contextual information
(Kersten-Oertel et al., 2016)	System designed to support the planning of brain tumour surgery by overlaying personalised and real-time updated brain tumour data on the patient himself/herself	Infrared tracking Optical tracking		Contextual information, X-ray vision
(Lee et al., 2016)	3D hologram-based prototype for training mental rotation in older adults	Leap motions		3D models and animation
(Kwon et al., 2014)	For training basic life support without an instructor. Consisted of simulated realistic scenario with visual and auditory feedback	Interactive mannequin and projector		X-ray vision, Contextual information
(Freschi et al., 2015)	Prototype for training ultrasound imaging by improving hand-eye co-ordination and 3D depth perception	Ultrasound, 6 DOF electromagnetic sensor		Interactive virtual objects, X-ray vision with overlaid ultrasound reading
(Buń et al., 2015)	Simulation of ultrasound imaging with a mannequin focusing on orientation of the ultrasound device and tracking the path of the motion	Ultrasound, Motion sensor		X-ray vision
(Condino et al., 2016)	Provided tactile feedback for arteries using a wearable pressure-sensitive glove. Hints and related information were also provided	Potentiometer		Haptic feedback, X-ray vision
(Cirulis and Liepina, 2014)	Interactive augmented 3D environment that provided passive tactile feedback for teaching veterinary anatomy skills by overlaying 3D augmentations on a physical model			3D models, X-ray vision
(Islam et al., 2016)	Video-based skill assessment of minimally invasive surgical skills by measuring hand-eye coordination and depth perception. It also provided real-time feedback and the score	Depth sensor camera		X-ray vision, Contextual information
(Koreeda et al., 2016)	Supporting system to allow the surgeons to visualise obscured regions in real time	2 endoscopic cameras		X-ray vision
(Shekhar et al., 2010)	A stereoscopic vision system for visualising surgical anatomy with true depth for training spatial perception	Ultrasound, scanner and optical trackers		X-ray vision
(Kamphuis et al., 2014)	Displayed internal organs in a real-time tracked body for anatomy education	kinect		X-ray vision, 3D model and animation
(Western et al., 2015)	System for ultrasound guided radiation therapy for precise penetration of the radiation to aim at the correct tissue	Ultrasound		X-ray vision
(Chong et al., 2015)	Training environment for manual ultrasonic testing using real-time visualisation of ultrasound data	Ultrasound		X-ray vision
(Allain et al., 2015)	Navigation system for visually-impaired children based on a game which required them to find the path using auditory cues	Oculus rift		Cues and clues

Continuation of Table 1

Studies	Description	Sensor Used	Experts Direct Involvement	Instructional Methods
(Chang et al., 2015)	Prototype for training the correct reactions in patients with cognitive impairments. The patients were provided with visual and auditory cues to support them to make the correct decision	Optical tracker		Cues and clues
(Sano et al., 2016)	System for supporting novice players to plan their actions to reduce the skill gap by visualising the ball trajectory in a collaborative soccer game	22 motion cameras and projectors to project the path of the ball		Contextual information
(Park et al., 2016)	A serious game platform to help children with ADHD using a dynamically adaptive intervention scheme where the character is controlled by brain-computer interface	EEG, Kinect		Feedback, Contextual information
(Zhu et al., 2014)	A maintenance support system that provided instructions and information about the equipment according to the expertise level of the user	Optical tracking, camera		Contextual information
(Dalle Mura et al., 2016)	Support system for manual assembly workers which superimposed the correct assembly sequence and provided feedback based on data from a pressure sensor placed under the box containing each part	Potentiometer		Contextual information, Feedback
(Sousa et al., 2016)	System to assist novice pool players using AR by projecting the shot path calculated based on the cue orientation of the player	Kinect projector		Contextual information
(Kim and Dey, 2016)	Prototype with the concept of augmenting drivers' senses to represent information in a way that minimised perceptual and cognitive load processing	Eye tracking EEG, heart rate sensor	Task analysis with expert to identify relevant aspects of the task	Haptic feedback, Contextual information
(Choi et al., 2017)	Surgical navigation system for bone tumour resection by augmenting the path of the laser beam	Ultrasound		X-ray vision
(Oyekan et al., 2017)	Remote collaboration platform with digitised human and workpiece for interactions allowing remote communication and interaction with the environment	Kinect		Interactive virtual objects, annotation
(Altimira et al., 2017)	Platform for enhancing table tennis player motivation by balancing gameplay digitally using a project or to highlight areas of the table where the expert may play, limiting his/her performance	Projector		Contextual information
(Such et al., 2014)	Prototype for training composite layup by guiding the trainee with instructions and feedback based on his/her hand movement	Kinect		Contextual information
(Sebillo et al., 2015)	Designed for training for collaboration in emergency situations, providing relevant information such as status of others and location data	GPS	The experts designed the simulation	Contextual information
(Borges et al., 2016)	Multimodal training system for electric-powered wheelchair users with brain-computer interface	EEG		Contextual information
(Perlini et al., 2014)	Guided cardiac auscultation training on a mannequin to assist residents to identify different cardiac diagnoses by listening to the heart murmurs			Contextual information, Object enrichment

Continuation of Table 1

Studies	Description	Sensor Used	Experts Direct Involvement	Instructional Methods
(Asadipour et al., 2017)	Provided multimodal visualisation to assist in palpation task. Also used gamification to improve skill acquisition	Potentiometer	Captured expert pressure application	Haptic feedback
(Matassa and Morreale, 2016)	Supporting singers with tactile and visual feedback on their performance in real time	Heart Rate sensor, Vibration collector		Feedback
(Funk et al., 2016)	Providing haptic auditory or visual feedback on Lego assembly task to measure the effectiveness of each type of feedback	Kinect		Haptic feedback
(Zhao et al., 2016)	Designed for the assessment of thumb release in archery and provided summative feedback on performance	Accelerometer infused in glove	Captured 50 samples from experts to use as a benchmark for the assessment	Feedback
(Ahmmad et al., 2014)	Designed to measure the psychomotor skills and dexterity of the novice for surgical procedures using a haptic-enabled surgical device	Haptic device	Captured expert performance on psychomotor skills	Haptic feedback
(Araki et al., 2017)	Prototype to assess the gripping and reaction force of the novice during laparoscopic surgery. The experiment was performed on a pig carcass	Pressure-sensitive forceps, Accelerometer	Recorded amount of pressure used by experts during the procedure	Haptic feedback
(Prabhu et al., 2017)	Training platform for composite layup task, providing feedback based on decoded key steps of the captured performance	Kinect	Captured expert performance and manually decoded the recording into individual steps	Feedback
(Horeman et al., 2014)	Provided visual feedback on the force applied on a knot-tying surgical procedure with different coloured lights indicating different levels	Potentiometer		Feedback
(Kowalewski et al., 2017)	Laparoscopic surgery training prototype based on an expert model which provided real-time feedback on the performance of the trainee in a simulated environment	Kinect, infrared cameras	Data collected from four experts' demonstration	Interactive virtual objects, Feedback

End of Table

Summary

This dissertation reports on the research conducted with the intention to investigate and support deliberate practice of complex skills using multimodal technologies. Complex skills are valuable human resources and are difficult to learn with an estimated five hundred hours to acquire proficiency. Multimodal technologies such as augmented reality, enable immersive authentic practice of complex skills. The research begins by studying the state-of-the-art on the use of multimodal technologies for training complex skills. Moving forward, following a design-based approach, the research conducts several studies in the context of the WEKIT project and also in the context of training calligraphy.

Exploration of the state-of-the-art on using multimodal technologies for training complex skill was done by conducting a systematic literature review (chapter 2) study which analyses 78 studies. The study extracts the instructional design patterns from these studies and analyses them according to the four components instructional design model (4C/ID) providing an overview of the potential of such design patterns to train complex skills with deliberate practice.

Based on the findings of the literature study, the "Instructional design for augmented reality" (ID4AR) framework was conceptualised (chapter 3). The framework provides a taxonomy of instructional design patterns and guidelines for implementing the framework to instructional designer. Further the framework also classifies the design patterns according to the task type, as defined by the original authors, in order to support the selection of design patterns ideal for the type of task that the instructional designer wishes to train. The framework was then, utilised to design two prototypes, namely WEKIT.one in the context of the WEKIT project and the calligraphy tutor.

The first user study using WEKIT.one was conducted in the three domains of the WEKIT project (chapter 4). This user study aimed to evaluate the WEKIT.one adherence to the ID4AR framework. The ID4AR framework and its design patterns are abstract from the domain and can be implemented in different ways depending on the context. In this study, the participants from the three WEKIT domains used the application to verify that the implemented design patterns indeed met their intended definition. Additionally, the usability of the WEKIT.one was also tested and found to be acceptable. Similarly, results from the study show that participants from the all three domains found that the WEKIT.one met the basic assumptions of the ID4AR framework.

After confirming that the WEKIT.One met the assumptions of the ID4AR framework, a second study was conducted in the context of the WEKIT project to evaluate its training efficacy (chapter 5). The WEKIT.One prototype allows the expert/mentor to create an expert model for training students with the help of sensors. This model is the shared and loaded into other installations of the prototype

by the students for training. In this study, first, an expert model was created using the WEKIT.one. This expert model was then evaluated by other experts in the domain to assess its suitability to be used for training. The results of this part of the study showed that experts in the three WEKIT domains agree that the expert model recorded with the WEKIT.One is suitable to be used for training. At the same time, the students' participated in the study in, either the control group using paper-based instructions or, the treatment group using the WEKIT.one prototype. The results of the post-test showed no difference in the performance scores between the groups.

A critical limitation that may be have been responsible for the results in the above-mentioned study could be the lack of comprehensive immediate feedback by the WEKIT.one prototype. Feedback plays a critical role in learning of a complex skill and also to foster deliberate practice. To explore how feedback should be designed in multimodal systems for training complex skill, the Calligraphy tutor was developed using the ID4AR framework (chapter 6). Doing so allowed the exploration of the framework in a new domain with fine motor skills and also the opportunity to develop a specific prototype to train a specific complex skill, both of which were in contrast to the WEKIT project. The calligraphy tutor provides feedback using various modalities on different aspects of the complex skills. The study compares the mental effort imposed by the feedback individually or together with the control group that received no feedback. Generally, there was no significant difference in the mental effort between the control and the treatment group. Similarly, there was no significant difference in the reported mental effort between the different modalities, in the treatment group. In addition, the usability of the calligraphy trainer was found to be acceptable.

The last chapter (chapter 7) discusses the main findings of the studies conducted in this thesis. These findings highlight the potential of multimodal technologies to support deliberate practice of complex skills. Moreover, the results also indicate that various types of skills, for e.g. psychomotor skills, can be trained with multimodal technologies. The ID4AR framework provides a systematic approach to implement the design patterns according to the 4C/ID model, thus supporting deliberate practice of complex skills using multimodal technologies. The discussion continues with the limitations of the research. These include the limitations such as those, imposed by the current state of technology and the WEKIT project. The thesis concludes by suggesting future research paths based on the findings of the previous studies conducted in this research.

Samenvetting

In dit proefschrift wordt verslag gedaan van het onderzoek dat is uitgevoerd met de bedoeling de bekende praktijk van complexe vaardigheden met behulp van multimodale technologieën te onderzoeken en te ondersteunen. Complexe vaardigheden zijn waardevolle menselijke hulpbronnen en deze zijn moeilijk te leren vanwege de geschatte vijfhonderd uur om deze vaardigheid te verwerven. Multimodale technologieën, zoals augmented reality, maken een immersieve authentieke praktijk van complexe vaardigheden mogelijk. Het onderzoek begint met het bestuderen van de state-of-the-art over het gebruik van multimodale technologieën voor het trainen van complexe vaardigheden. In het kader van het WEKIT-project en in het kader van de opleiding kalligrafie worden verschillende studies uitgevoerd.

Het onderzoek naar de stand van zaken met betrekking tot het gebruik van multimodale technologieën voor de opleiding van complexe vaardigheden is uitgevoerd door middel van een systematisch literatuuronderzoek (hoofdstuk 2), waarin 78 studies worden geanalyseerd. De studie haalt de instructieve ontwerppatronen uit deze studies en analyseert ze volgens het vier componenten tellende instructieve ontwerpmodel (4C/ID) dat een overzicht geeft van de mogelijkheden van dergelijke ontwerppatronen om complexe vaardigheden met een bekende praktijk te trainen.

Op basis van de bevindingen van de literatuurstudie is het "Instructional design for augmented reality" (ID4AR) raamwerk geconceptualiseerd (hoofdstuk 3). Het raamwerk biedt een taxonomie van instructieve ontwerppatronen en richtlijnen voor de implementatie van het raamwerk naar instructief ontwerper. Verder classificeert het kader ook de ontwerppatronen volgens het taaktype, zoals gedefinieerd door de oorspronkelijke auteurs. Dit ter ondersteuning van de selectie van ontwerppatronen die ideaal zijn voor het type taak dat de instructieontwerper wil opleiden. Het kader is vervolgens gebruikt om twee prototypes te ontwerpen, namelijk WEKIT.one in het kader van het WEKIT-project en de kalligrafie-tutor.

De eerste gebruikersstudie met behulp van WEKIT.one is uitgevoerd in de drie domeinen van het WEKIT project (hoofdstuk 4). Deze gebruikersstudie had als doel het WEKIT.one-schema te evalueren en de naleving van het ID4AR-raamwerk te evalueren. Het ID4AR-raamwerk en zijn ontwerppatronen zijn abstract van het domein en kunnen, afhankelijk van de context, op verschillende manieren worden geïmplementeerd. In deze studie hebben de deelnemers van de drie WEKIT-domeinen de applicatie gebruikt om na te gaan of de geïmplementeerde ontwerppatronen inderdaad voldoen aan de beoogde definitie. Daarnaast is ook de bruikbaarheid van het WEKIT.one getest en acceptabel bevonden. Evenzo blijkt uit de resultaten van het onderzoek dat deelnemers uit de drie domeinen vonden dat het WEKIT.one voldeed aan de basisveronderstellingen van het ID4AR-raamwerk.

Na de bevestiging dat het WEKIT.one voldeed aan de aannames van het ID4AR-raamwerk, is in het kader van het WEKIT-project een tweede studie uitgevoerd om de effectiviteit van de training te evalueren (hoofdstuk 5). Het WEKIT.1 prototype stelt de expert/mentor in staat om met behulp van sensoren een expertmodel te creëren voor het trainen van studenten. Dit model is gedeeld en geüpload in andere installaties van het prototype door de cursisten voor de opleiding. In deze studie is eerst een expertmodel gemaakt met behulp van het WEKIT.one. Dit expertmodel is vervolgens geëvalueerd door andere experts in het domein om te beoordelen of het geschikt is om te gebruiken voor training. De resultaten van dit deel van het onderzoek laten zien dat experts in de drie WEKIT-domeinen het erover eens zijn dat het expertmodel dat met het WEKIT is vastgelegd, geschikt is om te worden gebruikt voor training. Tegelijkertijd namen de studenten deel aan het onderzoek, hetzij in de controlegroep met behulp van papieren instructies, hetzij in de behandelgroep met behulp van het WEKIT.one prototype. De resultaten van de post-test lieten geen verschil zien in de prestatiescores tussen de groepen.

Een kritische beperking die verantwoordelijk kan zijn geweest voor de resultaten van bovengenoemd onderzoek, is het ontbreken van een uitgebreide directe terugkoppeling door het WEKIT.one prototype. Feedback speelt een kritische rol bij het aanleren van een complexe vaardigheid en ook bij het stimuleren van doelbewuste oefening. Om te onderzoeken hoe feedback moet worden ontworpen in multimodale systemen voor het trainen van complexe vaardigheden, is de kalligrafie-tutor ontwikkeld met behulp van het ID4AR-raamwerk (hoofdstuk 6). Hierdoor werd het mogelijk om het raamwerk te verkennen in een nieuw domein met fijne motorische vaardigheden en ook om een specifiek prototype te ontwikkelen om een specifieke complexe vaardigheid te trainen, beide in tegenstelling tot het WEKIT-project. De kalligraafdocent geeft feedback met behulp van verschillende manieren over verschillende aspecten van de complexe vaardigheden. Het onderzoek vergelijkt de mentale inspanning die de feedback oplegt, individueel of samen met de controlegroep die geen feedback kreeg. Over het algemeen was er geen significant verschil in de mentale inspanning tussen de controlegroep en de behandelgroep. Ook was er geen significant verschil in de gerapporteerde mentale inspanning tussen de verschillende modaliteiten in de behandelgroep. Daarnaast werd de bruikbaarheid van de kalligraaftrainer acceptabel bevonden.

In het laatste hoofdstuk (hoofdstuk 7) worden de belangrijkste bevindingen van de in dit proefschrift uitgevoerde studies besproken. Deze bevindingen benadrukken het potentieel van multimodale technologieën om de bewuste praktijk van complexe vaardigheden te ondersteunen. Bovendien geven de resultaten ook aan dat verschillende soorten vaardigheden, voor bijvoorbeeld psychomotorische vaardigheden, getraind kunnen worden met multimodale technologieën. Het ID4AR-raamwerk biedt een systematische aanpak om de ontwerppatronen volgens het 4C/ID-model te implementeren en ondersteunt zo de bewuste praktijk van complexe vaardigheden met behulp van multimodale technologieën. De discussie gaat verder met de beperkingen van het onderzoek. Het gaat onder meer om de beperkingen die de huidige stand van de techniek en het WEKIT-project met zich meebrengen. Het proefschrift sluit af met het voorstellen van toekomstige onderzoekstrajecten op basis van de bevindingen van de eerdere studies die in dit onderzoek zijn uitgevoerd.

SIKS Dissertation Series

The complete list of dissertations carried out under the auspices of SIKS, the Dutch Research School for Information and Knowledge Systems from 1998 on can be found at `http://www.siks.nl/dissertations.php`.

2011

2011-01 Botond Cseke (RUN)
Variational Algorithms for Bayesian Inference in Latent Gaussian Models.

2011-02 Nick Tinnemeier(UU)
Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language.

2011-03 Jan Martijn van der Werf (TUE)
Compositional Design and Verification of Component-Based Information Systems.

2011-04 Hado Philip van Hasselt (UU)
Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference learning algorithms.

2011-05 Bas van de Raadt (VU)
Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.

2011-06 Yiwen Wang(TUE)
Semantically-Enhanced Recommendations in Cultural Heritage.

2011-07 Yujia Cao (UT
Multimodal Information Presentation for High Load Human Computer Interaction.

2011-08 Nieske Vergunst (UU)
BDI-based Generation of Robust Task-Oriented Dialogues.

2011-09 Tim de Jong (OU)
Contextualised Mobile Media for Learning.

2011-10 Bart Bogaert (UVT)
Cloud Content Contention.

2011-11 Dhaval Vyas (UT)
Designing for Awareness: An Experience-focused HCI Perspective.

2011-12 Carmen Bratosin (TUE)
Grid Architecture for Distributed Process Mining.

2011-13 Xiaoyu Mao (UVT)
Airport under Control; Multiagent Scheduling for Airport Ground Handling.

2011-14 Milan Lovric(EUR)
Behavioral Finance and Agent-Based Artificial Markets.

2011-15 Marijn Koolen (UVA)
The Meaning of Structure: the Value of Link Evidence for Information Retrieval.

2011-16 Maarten Schadd (UM)
Selective Search in Games of Different Complexity.

2011-17 Jiyin He (UVA)
Exploring Topic Structure: Coherence, Diversity and Relatedness.

2011-18 Mark Ponsen (UM)
Strategic Decision-Making in complex games.

2011-19 Ellen Rusman (OU)
The Mind ’ s Eye on Personal Profiles.

2011-20 Qing Gu (VU)
Guiding service-oriented software engineering - A view-based approach.

2011-21 Linda Terlouw (TUD)
Modularization and Specification of Service-Oriented Systems.

2011-22 Junte Zhang (UVA)
System Evaluation of Archival Description and Access.

2011-23 Wouter Weerkamp (UVA)
Finding People and their Utterances in Social Media.

2011-24 Herwin van Welbergen (UT)
Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior.

2011-25 Syed Waqar ul Qounain Jaffry (VU)
Analysis and Validation of Models for Trust Dynamics.

2011-26 Matthijs Aart Pontier (VU)
Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots.

2011-27 Aniel Bhulai (VU)
Dynamic website optimization through autonomous management of design patterns.

2011-28 Rianne Kaptein (UVA)
Effective Focused Retrieval by Exploiting Query Context and Document Structure.

2011-29 Faisal Kamiran (TUE)
Discrimination-aware Classification.

2011-30 Egon van den Broek (UT)
Affective Signal Processing (ASP): Unraveling the mystery of emotions.

2011-31 Ludo Waltman (EUR)
Computational and Game-Theoretic Approaches for Modeling Bounded Rationality.

2011-32 Nees-Jan van Eck (EUR)
Methodological Advances in Bibliometric Mapping of Science.

2011-33 Tom van der Weide (UU)
Arguing to Motivate Decisions.

2011-34 Paolo Turrini (UU)
Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations.

2011-35 Maaike Harbers (UU)
Explaining Agent Behavior in Virtual Training.

2011-36 Erik van der Spek (UU)
Experiments in serious game design: a cognitive approach.

2011-37 Adriana Burlutiu (RUN)
Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference.

2011-38 Nyree Lemmens (UM)
Bee-inspired Distributed Optimization.

2011-39 Joost Westra (UU)
Organizing Adaptation using Agents in Serious Games.

2011-40 Viktor Clerc (VU)
Architectural Knowledge Management in Global Software Development.

2011-41 Luan Ibraimi (UT)
Cryptographically Enforced Distributed Data Access Control.

2011-42 Michal Sindlar (UU)
Explaining Behavior through Mental State Attribution.

2011-43 Henk van der Schuur (UU)
Process Improvement through Software Operation Knowledge.

2011-44 Boris Reuderink (UT)
Robust Brain-Computer Interfaces.

2011-45 Herman Stehouwer (UVT)
Statistical Language Models for Alternative Sequence Selection.

2011-46 Beibei Hu (TUD)
Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work.

2011-47 Azizi Bin Ab Aziz(VU)
Exploring Computational Models for Intelligent Support of Persons with Depression.

2011-48 Mark Ter Maat (UT)
Response Selection and Turn-taking for a Sensitive Artificial Listening Agent.

2011-49 Andreea Niculescu (UT)
Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality.

2012

2012-01 Terry Kakeeto (UVT)
Relationship Marketing for SMEs in Uganda.

2012-02 Muhammad Umair(VU)
Adaptivity, emotion, and Rationality in Human and Ambient Agent Models.

2012-03 Adam Vanya (VU)
Supporting Architecture Evolution by Mining Software Repositories.

2012-04 Jurriaan Souer (UU)
Development of Content Management System-based Web Applications.

2012-05 Marijn Plomp (UU)
Maturing Interorganisational Information Systems.

2012-06 Wolfgang Reinhardt (OU)
Awareness Support for Knowledge Workers in Research Networks.

2012-07 Rianne van Lambalgen (VU)
When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions.

2012-08 Gerben de Vries (UVA)
Kernel Methods for Vessel Trajectories.

2012-09 Ricardo Neisse (UT)
Trust and Privacy Management Support for Context-Aware Service Platforms.

2012-10 David Smits (TUE)
Towards a Generic Distributed Adaptive Hypermedia Environment.

2012-11 J.C.B. Rantham Prabhakara (TUE)
Process Mining in the Large: Preprocessing, Discovery, and Diagnostics.

2012-12 Kees van der Sluijs (TUE)
Model Driven Design and Data Integration in Semantic Web Information Systems.

2012-13 Suleman Shahid (UVT)
Fun and Face: Exploring non-verbal expressions of emotion during playful interactions.

2012-14 Evgeny Knutov(TUE)
Generic Adaptation Framework for Unifying Adaptive Web-based Systems.

2012-15 Natalie van der Wal (VU)
Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes..

2012-16 Fiemke Both (VU)
Helping people by understanding them - Ambient Agents supporting task execution and depression treatment.

2012-17 Amal Elgammal (UVT)
Towards a Comprehensive Framework for Business Process Compliance.

2012-18 Eltjo Poort (VU)
Improving Solution Architecting Practices.

2012-19 Helen Schonenberg (TUE)
What's Next? Operational Support for Business Process Execution.

2012-20 Ali Bahramisharif (RUN)
Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing.

2012-21 Roberto Cornacchia (TUD)
Querying Sparse Matrices for Information Retrieval.

2012-22 Thijs Vis (UVT)
Intelligence, politie en veiligheidsdienst: verenigbare grootheden?.

2012-23 Christian Muehl (UT)
Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction.

2012-24 Laurens van der Werff (UT)
Evaluation of Noisy Transcripts for Spoken Document Retrieval.

2012-25 Silja Eckartz (UT)
Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application.

2012-26 Emile de Maat (UVA)
Making Sense of Legal Text.

2012-27 Hayrettin Gurkok (UT
Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games.

2012-28 Nancy Pascall (UVT)
Engendering Technology Empowering Women.

2012-29 Almer Tigelaar (UT)
Peer-to-Peer Information Retrieval.

2012-30 Alina Pommeranz (TUD)
Designing Human-Centered Systems for Reflective Decision Making.

2012-31 Emily Bagarukayo (RUN)
A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure.

2012-32 Wietske Visser (TUD)
Qualitative multi-criteria preference representation and reasoning.

2012-33 Rory Sie (OUN)
Coalitions in Cooperation Networks (COCOON).

2012-34 Pavol Jancura (RUN)
Evolutionary analysis in PPI networks and applications.

2012-35 Evert Haasdijk (VU)
Never Too Old To Learn – On-line Evolution of Controllers in Swarm- and Modular Robotics.

2012-36 Denis Ssebugwawo (RUN)
Analysis and Evaluation of Collaborative Modeling Processes.

2012-37 Agnes Nakakawa (RUN)
A Collaboration Process for Enterprise Architecture Creation.

2012-38 Selmar Smit (VU)
Parameter Tuning and Scientific Testing in Evolutionary Algorithms.

2012-39 Hassan Fatemi (UT)
Risk-aware design of value and coordination networks.

2012-40 Agus Gunawan (UVT)
Information Access for SMEs in Indonesia.

2012-41 Sebastian Kelle (OU)
Game Design Patterns for Learning.

2012-42 Dominique Verpoorten (OU)
Reflection Amplifiers in self-regulated Learning.

2012-43 Withdrawn
.

2012-44 Anna Tordai (VU)
On Combining Alignment Techniques.

2012-45 Benedikt Kratz (UVT)
A Model and Language for Business-aware Transactions.

2012-46 Simon Carter (UVA)
Exploration and Exploitation of Multilingual Data for Statistical Machine Translation.

2012-47 Manos Tsagkias (UVA)
Mining Social Media: Tracking Content and Predicting Behavior.

2012-48 Jorn Bakker (TUE)
Handling Abrupt Changes in Evolving Timeseries Data.

2012-49 Michael Kaisers (UM)
Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions.

2012-50 Steven van Kervel (TUD)
Ontology driven Enterprise Information Systems Engineering.

2012-51 Jeroen de Jong (TUD)
Heuristics in Dynamic Scheduling; a practical framework with a case study in elevator dispatching.

2013

2013-01 Viorel Milea (EUR)
News Analytics for Financial Decision Support.

2013-02 Erietta Liarou (CWI)
MonetDB/DataCell: Leveraging the Columnstore Database Technology for Efficient and Scalable Stream Processing.

2013-03 Szymon Klarman (VU)
Reasoning with Contexts in Description Logics.

2013-04 Chetan Yadati(TUD)
Coordinating autonomous planning and scheduling.

2013-05 Dulce Pumareja (UT)
Groupware Requirements Evolutions Patterns.

2013-06 Romulo Goncalves(CWI)
The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience.

2013-07 Giel van Lankveld (UVT)
Quantifying Individual Player Differences.

2013-08 Robbert-Jan Merk(VU)

Making enemies: cognitive modeling for opponent agents in fighter pilot simulators.

2013-09 Fabio Gori (RUN)
Metagenomic Data Analysis: Computational Methods and Applications.

2013-10 Jeewanie Jayasinghe Arachchige(UVT)
A Unified Modeling Framework for Service Design..

2013-11 Evangelos Pournaras(TUD)
Multi-level Reconfigurable Self-organization in Overlay Services.

2013-12 Maryam Razavian(VU)
Knowledge-driven Migration to Services.

2013-13 Mohammad Safiri(UT)
Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly.

2013-14 Jafar Tanha (UVA)
Ensemble Approaches to Semi-Supervised Learning Learning.

2013-15 Daniel Hennes (UM)
Multiagent Learning - Dynamic Games and Applications.

2013-16 Eric Kok (UU)
Exploring the practical benefits of argumentation in multi-agent deliberation.

2013-17 Koen Kok (VU)
The PowerMatcher: Smart Coordination for the Smart Electricity Grid.

2013-18 Jeroen Janssens (UVT)
"Outlier Selection and One-Class Classification".

2013-19 Renze Steenhuizen (TUD)
Coordinated Multi-Agent Planning and Scheduling.

2013-20 Katja Hofmann (UVA)
Fast and Reliable Online Learning to Rank for Information Retrieval.

2013-21 Sander Wubben (UVT)
Text-to-text generation by monolingual machine translation.

2013-22 Tom Claassen (RUN)
Causal Discovery and Logic.

2013-23 Patricio de Alencar Silva(UVT)
Value Activity Monitoring.

2013-24 Haitham Bou Ammar (UM)
Automated Transfer in Reinforcement Learning.

2013-25 Agnieszka Anna Latoszek-Berendsen (UM)
Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System.

2013-26 Alireza Zarghami (UT)
Architectural Support for Dynamic Homecare Service Provisioning.

2013-27 Mohammad Huq (UT)
Inference-based Framework Managing Data Provenance.

2013-28 Frans van der Sluis (UT)
When Complexity becomes Interesting: An Inquiry into the Information eXperience.

2013-29 Iwan de Kok (UT)
Listening Heads.

2013-30 Joyce Nakatumba (TUE)
Resource-Aware Business Process Management: Analysis and Support.

2013-31 Dinh Khoa Nguyen (UVT)
Blueprint Model and Language for Engineering Cloud Applications.

2013-32 Kamakshi Rajagopal (OUN)
Networking For Learning; The role of Networking in a Lifelong Learner's Professional Development.

2013-33 Qi Gao (TUD)
User Modeling and Personalization in the Microblogging Sphere.

2013-34 Kien Tjin-Kam-Jet (UT)
Distributed Deep Web Search.

2013-35 Abdallah El Ali (UVA)
Minimal Mobile Human Computer Interaction.

2013-36 Than Lam Hoang (TUe)
Pattern Mining in Data Streams.

2013-37 Dirk Borner (OUN)
Ambient Learning Displays.

2013-38 Eelco den Heijer (VU)
Autonomous Evolutionary Art.

2013-39 Joop de Jong (TUD)

A Method for Enterprise Ontology based Design of Enterprise Information Systems.

2013-40 Pim Nijssen (UM)
Monte-Carlo Tree Search for Multi-Player Games.

2013-41 Jochem Liem (UVA)
Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning.

2013-42 Leon Planken (TUD)
Algorithms for Simple Temporal Reasoning.

2013-43 Marc Bron (UVA)
Exploration and Contextualization through Interaction and Concepts.

2014

2014-01 Nicola Barile (UU)
Studies in Learning Monotone Models from Data.

2014-02 Fiona Tuliyano (RUN)
Combining System Dynamics with a Domain Modeling Method.

2014-03 Sergio Raul Duarte Torres (UT)
Information Retrieval for Children: Search Behavior and Solutions.

2014-04 Hanna Jochmann-Mannak (UT)
Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation.

2014-05 Jurriaan van Reijsen (UU)
Knowledge Perspectives on Advancing Dynamic Capability.

2014-06 Damian Tamburri (VU)
Supporting Networked Software Development.

2014-07 Arya Adriansyah (TUE)
Aligning Observed and Modeled Behavior.

2014-08 Samur Araujo (TUD)
Data Integration over Distributed and Heterogeneous Data Endpoints.

2014-09 Philip Jackson (UVT)
Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language.

2014-10 Ivan Salvador Razo Zapata (VU)
Service Value Networks.

2014-11 Janneke van der Zwaan (TUD)
An Empathic Virtual Buddy for Social Support.

2014-12 Willem van Willigen (VU)
Look Ma, No Hands: Aspects of Autonomous Vehicle Control.

2014-13 Arlette van Wissen (VU)
Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains.

2014-14 Yangyang Shi (TUD)
Language Models With Meta-information.

2014-15 Natalya Mogles (VU)
Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare.

2014-16 Krystyna Milian (VU)
Supporting trial recruitment and design by automatically interpreting eligibility criteria.

2014-17 Kathrin Dentler (VU)
Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability.

2014-18 Mattijs Ghijsen (UVA)
Methods and Models for the Design and Study of Dynamic Agent Organizations.

2014-19 Vinicius Ramos (TUE)
Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support.

2014-20 Mena Habib (UT)
Named Entity Extraction and Disambiguation for Informal Text: The Missing Link.

2014-21 Kassidy Clark (TUD)
Negotiation and Monitoring in Open Environments.

2014-22 Marieke Peeters (UU)
Personalized Educational Games - Developing agent-supported scenario-based training.

2014-23 Eleftherios Sidirourgos (UVA/CWI)
Space Efficient Indexes for the Big Data Era.

2014-24 Davide Ceolin (VU)

Trusting Semi-structured Web Data.

2014-25 Martijn Lappenschaar (RUN)
New network models for the analysis of disease interaction.

2014-26 Tim Baarslag (TUD)
What to Bid and When to Stop.

2014-27 Rui Jorge Almeida (EUR)
Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty.

2014-28 Anna Chmielowiec (VU)
Decentralized k-Clique Matching.

2014-29 Jaap Kabbedijk (UU)
Variability in Multi-Tenant Enterprise Software.

2014-30 Peter de Cock (UVT)
Anticipating Criminal Behaviour.

2014-31 Leo van Moergestel (UU)
Agent Technology in Agile Multiparallel Manufacturing and Product Support.

2014-32 Naser Ayat (UVA)
On Entity Resolution in Probabilistic Data.

2014-33 Tesfa Tegegne (RUN)
Service Discovery in eHealth.

2014-34 Christina Manteli(VU)
The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems..

2014-35 Joost van Oijen (UU)
Cognitive Agents in Virtual Worlds: A Middleware Design Approach.

2014-36 Joos Buijs (TUE)
Flexible Evolutionary Algorithms for Mining Structured Process Models..

2014-37 Maral Dadvar (UT)
Experts and Machines United Against Cyberbullying.

2014-38 Danny Plass-Oude Bos (UT)
Making brain-computer interfaces better: improving usability through post-processing..

2014-39 Jasmina Maric (UVT)
Web Communities, Immigration, and Social Capital.

2014-40 Walter Omona (RUN)
A Framework for Knowledge Management Using ICT in Higher Education..

2014-41 Frederic Hogenboom (EUR)
Automated Detection of Financial Events in News Text.

2014-42 Carsten Eijckhof (CWI/TUD)
Contextual Multidimensional Relevance Models.

2014-43 Kevin Vlaanderen (UU)
Supporting Process Improvement using Method Increments.

2014-44 Paulien Meesters (UVT)
Intelligent Blauw: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden..

2014-45 Birgit Schmitz (OUN)
Mobile Games for Learning: A Pattern-Based Approach.

2014-46 Ke Tao (TUD)
Social Web Data Analytics: Relevance, Redundancy, Diversity.

2014-47 Shangsong Liang (UVA)
Fusion and Diversification in Information Retrieval.

2015

2015-01 Niels Netten (UVA)
Machine Learning for Relevance of Information in Crisis Response.

2015-02 Faiza Bukhsh (UVT)
Smart auditing: Innovative Compliance Checking in Customs Controls.

2015-03 Twan van Laarhoven (RUN)
Machine learning for network data.

2015-04 Howard Spoelstra (OUN)
Collaborations in Open Learning Environments.

2015-05 Christoph Bosch (UT)
Cryptographically Enforced Search Pattern Hiding.

2015-06 Farideh Heidari (TUD)
Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes.

2015-07 Maria-Hendrike Peetz (UVA)
Time-Aware Online Reputation Analysis.

2015-08 Jie Jiang (TUD)
Organizational Compliance: An agent-based model for designing and evaluating organizational interactions.

2015-09 Randy Klaassen (UT)
HCI Perspectives on Behavior Change Support Systems.

2015-10 Henry Hermans (OUN)
OpenU: design of an integrated system to support lifelong learning.

2015-11 Yongming Luo (TUE)
Designing algorithms for big graph datasets: A study of computing bisimulation and joins.

2015-12 Julie M. Birkholz (VU)
Modi Operandi of Social Network Dynamics: The Effect of Context on Scientific Collaboration Networks Promotor: Prof. dr. P. Groenewegen (VU), Prof. dr. J.H. Akkermans (VU).

2015-13 Giuseppe Procaccianti (VU)
Energy-Efficient Software.

2015-14 Bart van Straalen (UT)
A cognitive approach to modeling bad news conversations.

2015-15 Klaas Andries de Graaf (VU)
Ontology-based Software Architecture Documentation.

2015-16 Changyun Wei (UT)
Cognitive Coordination for Cooperative Multi-Robot Teamwork.

2015-17 Andre van Cleeff (UT)
Physical and Digital Security Mechanisms: Properties, Combinations and Trade-offs.

2015-18 Holger Pirk (CWI)
Waste Not, Want Not! - Managing Relational Data in Asymmetric Memories.

2015-19 Bernardo Tabuenca (OUN)
Ubiquitous Technology for Lifelong Learners.

2015-20 Lois Vanhee (UU)
Using Culture and Values to Support Flexible Coordination Using Culture and Values to Support Flexible Coordination.

2015-21 Sibren Fetter (OUN)
Using Culture and Values to Support Flexible CoordinationUsing Peer-Support to Expand and Stabilize Online Learning.

2015-22 Zhemin Zhu (UT)
Co-occurrence Rate Networks - Towards separate training for undirected graphical models.

2015-23 Luit Gazendam (VU)
Using Culture and Values to Support Flexible CoordinationCataloguer Support in Cultural Heritage.

2015-24 Richard Berendsen (UVA)
Finding People, Papers, and Posts: Vertical Search Algorithms and Evaluation.

2015-25 Steven Woudenberg (UU)
Bayesian Tools for Early Disease Detection.

2015-26 Alexander Hogenboom (EUR)
Sentiment Analysis of Text Guided by Semantics and Structure.

2015-27 Sandor Heman (CWI)
Updating compressed colomn stores.

2015-28 Janet Bagorogoza (TiU)
Knowledge Management and High Performance; The Uganda Financial Institutions Model for HPO.

2015-29 Hendrik Baier (UM)
Monte-Carlo Tree Search Enhancements for One-Player and Two-Player Domains.

2015-30 Kiavash Bahreini (OU)
Real-time Multimodal Emotion Recognition in E-Learning.

2015-31 Yakup Koc (TUD)
On the robustness of Power Grids.

2015-32 Jerome Gard (UL)
Corporate Venture Management in SMEs.

2015-33 Frederik Schadd (UM)
Ontology Mapping with Auxiliary Resources.

2015-34 Victor de Graaff (UT)
Geosocial Recommender Systems.

2015-35 Jungxao Xu (TUD)
Affective Body Language of Humanoid Robots: Perception and Effects in Human Robot Interaction.

2016

2016-01 Syed Saiden Abbas (RUN)
Recognition of Shapes by Humans and Machines.

2016-015 Steffen Michels (RUN)
Hybrid Probabilistic Logics - Theoretical Aspects, Algorithms and Experiments.

2016-017 Berend Weel (VU)
Towards Embodied Evolution of Robot Organisms.

2016-019 Julia Efremova (TUE)
Mining Social Structures from Genealogical Data.

2016-02 Michiel Meulendijk (UU)
Optimizing medication reviews through decision support: prescribing a better pill to swallow.

2016-03 Maya Sappelli (RUN)
Knowledge Work in Context: User Centered Knowledge Worker Support.

2016-04 Laurens Rietveld (VU)
Publishing and Consuming Linked Data.

2016-05 Evgeny Sherkhonov (UVA)
Expanded Acyclic Queries: Containment and an Application in Explaining Missing Answers.

2016-06 Michel Wilson (TUD)
Robust scheduling in an uncertain environment.

2016-07 Jeroen de Man (VU)
Measuring and modeling negative emotions for virtual training.

2016-08 Matje van de Camp (TiU)
A Link to the Past: Constructing Historical Social Networks from Unstructured Data.

2016-09 Archana Nottamkandath (VU)
Trusting Crowdsourced Information on Cultural Artefacts.

2016-10 George Karafotias (VU)
Parameter Control for Evolutionary Algorithms.

2016-11 Anne Schuth (UVA)
Search Engines that Learn from Their Users.

2016-12 Max Knobbout (UU)
Logics for Modelling and Verifying Normative Multi-Agent Systems.

2016-13 Nana Baah Gyan (VU)
The Web, Speech Technologies and Rural Development in West Africa - An ICT4D Approach.

2016-14 Ravi Khadka(UU)
Revisiting Legacy Software System Modernization.

2016-16 Guangliang Li (UVA)
Socially Intelligent Autonomous Agents that Learn from Human Reward.

2016-18 Albert Merono Penuela (VU)
Refining Statistical Data on the Web.

2016-20 Daan Odijk (VU)
Context & Semantics in News & Web Search.

2016-21 Alejandro Moreno Celleri (UT)
From Traditional to Interactive Playspaces: Automatic Analysis of Player Behavior in the Interactive Tag Playground.

2016-22 Grace Lewis (VU)
Software Architecture Strategies for Cyber-Foraging Systems.

2016-23 Fei Cai (UVA)
Query Auto Completion in Information Retrieval.

2016-24 Brend Wanders (UT)
Repurposing and Probabilistic Integration of Data; An Iterative and data model independent approach.

2016-25 Y. Kiseleva (TUE)
Using Contextual Information to Understand Searching and Browsing Behavior.

2016-26 Dilhan J. Thilakarathne (VU)
In or Out of Control: Exploring Computational Models to Study the Role of Human Awareness and Control in Behavioural Choices, with Applications in Aviation and Energy Management Domains.

2016-27 Wen Li (TUD)
Understanding Geo-spatial Information on Social Media.

2016-28 Mingxin Zhang (TUD)
Large-scale agent-based social simulation: A study on epidemic prediction and control.

2016-29 Nicolas Honing (TUD)
Understanding Geo-spatial Information on Social Media.

2016-30 Ruud Mattheij (UVT)
The Eyes Have IT.

2016-31 Mohammad Khelghati (UT)
Deep web content monitoring.

2016-32 Eelco Vriezekolk (UVT)
Assessing Telecommunication Service Availability Risks for Crisis Organisations.

2016-33 Peter Bloem (UVA)
Single Sample Statistics, exercises in learning from just one example.

2016-34 Dennis Schunselaar (TUE)
Configurable Process Trees: Elicitation, Analysis, and Enactment.

2016-35 Zhaochun Ren (UVA)
Monitoring Social Media: Summarization, Classification and Recommendation.

2016-36 Daphne Karreman (UT)
Beyond R2D2: The design of nonverbal interaction behavior optimized for robot-specific morphologies.

2016-37 Giovanni Sileno (UVA)
Aligning Law and Action - a conceptual and computational inquiry.

2016-38 Andrea Minuto (UT)
Materials that matter - Smart Materials meet Art & Interaction Design.

2016-39 Merijn Bruijnes
Believable Suspect Agents; Response and Interpersonal Style Selection for an Artificial Suspect.

2016-40 Christian Detweiler (TUD)
Accounting for Values in Design.

2016-41 Thomas King (TUD)
Governing Governance: A Formal Framework for Analysing Institutional Design and Enactment Governance.

2016-42 Spyros Martzoukos (UVA)
Combinatorial and compositional aspects of bilingual aligned corpora.

2016-43 Saskia Koldijk (RUN)
Context-Aware Support for Stress Self-Management: From Theory to Practice.

2016-44 Thibault Sellam (UVA)
Automatic assistants for database exploration.

2016-45 Bram van Laar (UT)
Experiencing Brain-Computer Interface Control.

2016-46 Jorge Gallego Perez (UT)
Robots to Make you Happy.

2016-47 Christina Weber (UL)
Real-time foresight - Preparedness for dynamic innovation networks.

2016-48 Tanja Buttler (TUD)
Collecting Lessons Learned.

2016-49 Gleb Polevoy (TUD)
Participation and Interaction in Projects: A Game-Theoretic Analysis.

2016-50 Yan Wang (UVT)
The Bridge of Dreams: Towards a Method for Operational Performance Alignment in IT-enabled Service Supply Chains.

2017

2017-01 Jan-Jaap Oerlemans (UL)
Investigating Cybercrime.

2017-02 Sjoerd Timmer (UU)
Designing and Understanding Forensic Bayesian Networks using Argumentation.

2017-03 Daniel Harold Telgen (UU)
Grid Manufacturing; A Cyber-Physical Approach with Autonomous Products and Reconfigurable Manufacturing Machines.

2017-04 Mrunal Gawade (CWI)
Multi-core parallelism in a column-store.

2017-05 Mahdieh Shadi (UVA)
Collaboration Behavior; Enhancement in Co-development.

2017-06 Damir Vandic (EUR)
Intelligent Information Systems for Web Product Search.

2017-07 Roel Bertens (UU)
Insight in Information: from Abstract to Anomaly.

2017-08 Rob Konijn (VU)
Detecting Interesting Differences:Data Mining in Health Insurance Data using Outlier Detec-

tion and Subgroup Discovery.

2017-09 Dong Nguyen (UT)
Text as Social and Cultural Data: A Computational Perspective on Variation in Text.

2017-10 Robby van Delden (UT)
(Steering) Interactive Play Behavior.

2017-11 Florian Kunneman (RUN)
Modelling patterns of time and emotion in Twitter #anticipointment.

2017-12 Sander Leemans (TUE)
Robust Process Mining with Guarantees.

2017-13 Gijs Huisman (UT)
Social Touch Technology - Extending the reach of social touch through haptic technology.

2017-14 Shoshannah Tekofsky (UVT)
You Are Who You Play You Are: Modelling Player Traits from Video Game Behavior.

2017-15 Peter Berck (RUN)
Memory-Based Text Correction.

2017-16 Aleksandr Chuklin (UVA)
Understanding and Modeling Users of Modern Search Engines.

2017-17 Daniel Dimov (UL)
Crowdsourced Online Dispute Resolution.

2017-18 Ridho Reinanda (UVA)
Entity Associations for Search.

2017-19 Jeroen Vuurens (TUD)
Proximity of Terms, Texts and Semantic Vectors in Information Retrieval.

2017-20 Mohammadbashir Sedighi (TUD)
Fostering Engagement in Knowledge Sharing: The Role of Perceived Benefits, Costs and Visibility.

2017-21 Jeroen Linssen (UT)
Meta Matters in Interactive Storytelling and Serious Gaming (A Play on Worlds).

2017-22 Sara Magliacane (VU)
Logics for causal inference under uncertainty.

2017-23 David Graus (UVA)
Entities of Interest — Discovery in Digital Traces.

2017-24 Chang Wang (TUD)
Use of Affordances for Efficient Robot Learning.

2017-25 Veruska Zamborlini (VUA)
Knowledge Representation for Clinical Guidelines, with applications to Multimorbidity Analysis and Literature Search.

2017-26 Merel Jung (UT)
Socially intelligent robots that understand and respond to human touch.

2017-27 Michiel Joosse (UT)
Investigating Positioning and Gaze Behaviors of Social Robots: People's Preferences, Perceptions and Behaviors.

2017-28 John Klein (VU)
Architecture Practices for Complex Contexts.

2017-29 Adel Alhuraibi (UVT)
From IT-BusinessStrategic Alignment to Performance: A Moderated Mediation Model of Social Innovation, and Enterprise Governance of IT.

2017-30 Wilma Latuny (UVT)
The Power of Facial Expressions.

2017-31 Ben Ruijl (UL)
Advances in computational methods for QFT calculations.

2017-32 Thaer Samar (RUN)
Access to and Retrievability of Content in Web Archives.

2017-33 Brigit van Loggem (OU)
Towards a Design Rationale for Software Documentation: A Model of Computer-Mediated Activity.

2017-34 Maren Scheffel (OUN)
The Evaluation Framework for Learning Analytics.

2017-35 Martine de Vos (VU)
Interpreting natural science spreadsheets.

2017-36 Yuanhao Guo (UL)
Shape Analysis for Phenotype Characterisation from High-throughput Imaging.

2017-37 Alejandro Montes Garcia (TUE)
WiBAF: A Within Browser Adaptation Framework that Enables Control over Privacy.

2017-38 Abdullah Kayal (TUD)
Normative Social Applications.

2017-39 Sara Ahmadi (RUN)
Exploiting properties of the human auditory system and compressive sensing methods to in-

crease noise robustness in ASR.

2017-40 Altaf Hussain Abro (VUA)
Steer your Mind: Computational Exploration of Human Control in Relation to Emotions, Desires and Social Support For applications in human-aware support systems".

2017-41 Adnan Manzoor (VUA)
Minding a Healthy Lifestyle:An Exploration of Mental Processes and a Smart Environment to Provide Support for a Healthy Lifestyle.

2017-42 Elena Sokolova (RUN)
Causal discovery from mixed and missing data with applications on ADHD datasets.

2017-43 Maaike de Boer (RUN)
Semantic Mapping in Video Retrieval.

2017-44 Garm Lucassen (UU)
Understanding User Stories - Computational Linguistics in Agile Requirements Engineering.

2017-45 Bas Testerink (UU)
Decentralized Runtime Norm Enforcement.

2017-46 Jan Schneider (OU)
Sensor-based Learning Support.

2017-47 Yie Yang (TUD)
Crowd Knowledge Creation Acceleration.

2017-48 Angel Suarez (OU)
Collaborative inquiry-based learning.

2018

2018-01 Han van der Aa (VU)
Comparing and Aligning Process Representations.

2018-02 Felix Mannhardt (TUE)
Multi-perspective Process Mining.

2018-03 Steven Bosems (UT)
Causal Models For Well-Being: Knowledge Modeling, Model-Driven Development of Context-Aware Applications, and Behavior Prediction.

2018-04 Jordan Janeiro (TUD)
Flexible Coordination Support for Diagnosis Teams in Data-Centric Engineering Tasks.

2018-05 Hugo Huurdeman (UVA)
Supporting the Complex Dynamics of the Information Seeking Process.

2018-06 Dan Ionita (UT)
Model-Driven Information Security Risk Assessment of Socio-Technical Systems.

2018-07 Jieting Luo (UU)
A formal account of opportunism in multi-agent systems.

2018-08 Rick Smetsers (RUN)
Advances in Model Learning for Software Systems.

2018-09 Xu Xie (TUD)
Data Assimilation in Discrete Event Simulations.

2018-10 Julienka Mollee (VUA)
Moving forward: supporting physical activity behavior change through intelligent technology.

2018-11 Mahdi Sargolzaei (UVA)
Enabling Framework for Service-oriented Collaborative Networks.

2018-12 Xixi Lu (TUE)
Using behavioral context in process mining.

2018-13 Seyed Amin Tabatabaei (VUA)
Computing a Sustainable Future: Exploring the added value of computational models for increasing the use of renewable energy in the residential sector.

2018-14 Bart Joosten (UVT)
Detecting Social Signals with Spatiotemporal Gabor Filters.

2018-15 Naser Davarzani (UM)
Biomarker discovery in heart failure.

2018-16 Jaebok Kim (UT)
Automatic recognition of engagement and emotion in a group of children.

2018-17 Jianpeng Zhang (TUE)
On Graph Sample Clustering.

2018-18 Henriette Nakad (UL)
De Notaris en Private Rechtspraak.

2018-19 Minh Duc Pham (VUA)
Emergent relational schemas for RDF.

2018-20 Manxia Liu (RUN)
Time and Bayesian Networks.

2018-21 Aad Slootmaker (OU)
EMERGO: a generic platform for authoring and playing scenario-based serious games.

2018-22 Eric Fernandes de Mello Araujo (VUA)
Contagious: Modeling the Spread of Behaviours, Perceptions and Emotions in Social Networks.

2018-23 Kim Schouten (EUR)
Semantics-driven Aspect-Based Sentiment Analysis.

2018-24 Jered Vroon (UT)
Responsive Social Positioning Behaviour for Semi-Autonomous Telepresence Robots.

2018-25 Riste Gligorov (VUA)
Serious Games in Audio-Visual Collections.

2018-26 Roelof de Vries (UT)
Theory-Based And Tailor-Made: Motivational Messages for Behavior Change Technology.

2018-27 Maikel Leemans (TUE)
Hierarchical Process Mining for Scalable Software Analysis.

2018-28 Christian Willemse (UT)
Social Touch Technologies: How they feel and how they make you feel.

2018-29 Yu Gu (UVT)
Emotion Recognition from Mandarin Speech.

2018-30 Wouter Beek (VU)
The K in semantic web stands for knowledge: scaling semantics to the web.

2019

2019-01 Rob van Eijk (UL)
Web privacy measurement in real-time bidding systems. A graph-based approach to RTB system classification.

2019-02 Emmanuelle Beauxis- Aussalet (CWI, UU)
Statistics and Visualizations for Assessing Class Size Uncertainty.

2019-03 Eduardo Gonzalez Lopez de Murillas (TUE)
Process Mining on Databases: Extracting Event Data from Real Life Data Sources.

2019-04 Ridho Rahmadi (RUN)
Finding stable causal structures from clinical data.

2019-05 Sebastiaan van Zelst (TUE)
Process Mining with Streaming Data.

2019-06 Chris Dijkshoorn (VU)
Nichesourcing for Improving Access to Linked Cultural Heritage Datasets.

2019-07 Soude Fazeli (TUD)
Recommender Systems in Social Learning Platforms.

2019-08 Frits de Nijs (TUD)
Resource-constrained Multi-agent Markov Decision Processes.

2019-09 Fahimeh Alizadeh Moghaddam (UVA)
Self-adaptation for energy efficiency in software systems.

2019-10 Qing Chuan Ye (EUR)
Multi-objective Optimization Methods for Allocation and Prediction.

2019-11 Yue Zhao (TUD)
Learning Analytics Technology to Understand Learner Behavioral Engagement in MOOCs.

2019-12 Jacqueline Heinerman (VU)
Better Together.

2019-13 Guanliang Chen (TUD)
MOOC Analytics: Learner Modeling and Content Generation.

2019-14 Daniel Davis (TUD)
Large-Scale Learning Analytics: Modeling Learner Behavior & Improving Learning Outcomes in Massive Open Online Courses.

2019-15 Erwin Walraven (TUD)
Planning under Uncertainty in Constrained and Partially Observable Environments.

2019-16 Guangming Li (TUE)
Process Mining based on Object-Centric Behavioral Constraint (OCBC) Models.

2019-17 Ali Hurriyetoglu (RUN)

Extracting actionable information from microtexts.

2019-18 Gerard Wagenaar (UU)
Artefacts in Agile Team Communication.

2019-19 Vincent Koeman (TUD)
Tools for Developing Cognitive Agents.

2019-20 Chide Groenouwe (UU)
Fostering technically augmented human collective intelligence.

2019-21 Cong Liu (TUE)
Software Data Analytics: Architectural Model Discovery and Design Pattern Detection.

2019-22 Martin van den Berg (VU)
Improving IT Decisions with Enterprise Architecture.

2019-23 Qin Lin (TUD)
Intelligent Control Systems: Learning, Interpreting, Verification.

2019-24 Anca Dumitrache (VU)
Truth in Disagreement- Crowdsourcing Labeled Data for Natural Language Processing.

2019-25 Emiel van Miltenburg (VU)
Pragmatic factors in (automatic) image description.

2019-26 Prince Singh (UT)
An Integration Platform for Synchromodal Transport.

2019-27 Alessandra Antonaci (OUN)
The Gamification Design Process applied to (Massive) Open Online Courses.

2019-28 Esther Kuindersma (UL)
Cleared for take-off:Game-based learning to prepare airline pilots for critical situations.

2019-29 Daniel Formolo (VU)
Using virtual agents for simulation and training of social skills in safety-critical circumstances.

2019-30 Vahid Yazdanpanah (UT)
Multiagent Industrial Symbiosis Systems.

2019-31 Milan Jelisavcic (VU)
Alive and Kicking: Baby Steps in Robotics.

2019-32 Chiara Sironi (UM)
Monte-Carlo Tree Search for Artificial General Intelligence in Games.

2019-33 Anil Yaman (TUE)
Evolution of Biologically Inspired Learning in Artificial Neural Networks.

2019-34 Negar Ahmadi (TUE)
EEG Microstate and Functional Brain Network Features for Classification of Epilepsy and PNES.

2019-35 Lisa Facey-Shaw (OUN)
Gamification with digital badges in learning programming.

2019-36 Kevin Ackermans (OUN)
Designing Video-Enhanced Rubrics to Master Complex Skills.

2019-37 Jian Fang (TUD)
Database Acceleration on FPGAs.

2019-38 Akos Kadar (OUN)
Learning visually grounded and multilingual representations.

2020

2020-01 Armon Toubman (UL)
Calculated Moves: Generating Air Combat Behaviour.

2020-02 Marcos de Paula Bueno (UL)
Unraveling Temporal Processes using Probabilistic Graphical Models.

2020-03 Mostafa Deghani (UvA)
Learning with Imperfect Supervision for Language Understanding.

2020-04 Maarten van Gompel (RUN)
Context as Linguistic Bridges.

2020-05 Yulong Pei (TUE)
On local and global structure mining.

2020-06 Preethu Rose Anish (UT)
Stimulation Architectural Thinking during Requirements Elicitation - An Approach and Tool Support.

2020-07 Wim van der Vegt (OUN)
Towards a software architecture for reusable game components.

2020-08 Ali Mirsoleimani (UL)
Structured Parallel Programming for Monte

Carlo Tree Search.

2020-09 Myriam Traub (UU)
Measuring Tool Bias & Improving Data Quality for Digital Humanities Research.

2020-10 Alifah Syamsiyah (TUE)
In-database Preprocessing for Process Mining.

2020-11 Sepideh Mesbah (TUD)
Semantic-Enhanced Training Data AugmentationMethods for Long-Tail Entity Recognition Models.

2020-12 Ward van Breda (VU)
Predictive Modeling in E-Mental Health: Exploring Applicability in Personalised Depression Treatment.

2020-13 Marco Virgolin (CWI/ TUD)
Design and Application of Gene-pool Optimal Mixing Evolutionary Algorithms for Genetic Programming.

2020-14 Mark Raasveldt (CWI/UL)
Integrating Analytics with Relational Databases.

2020-15 Georgiadis Konstantinos (OU)
Smart CAT: Machine Learning for Configurable Assessments in Serious Games.

2020-16 Ilona Wilmont (RUN)
Cognitive Aspects of Conceptual Modelling.

2020-17 Daniele Di Mitri (OU)
The Multimodal Tutor: Adaptive Feedback from Multimodal Experiences.

2020-18 Georgios Methenitis (TUD)
Agent Interactions & Mechanisms in Markets with Uncertainties: Electricity Markets in Renewable Energy Systems.

2020-19 Guido van Capelleveen (UT)
Industrial Symbiosis Recommender Systems.

2020-20 Albert Hankel (VU)
Embedding Green ICT Maturity in Organisations.

2020-21 Karine da Silva Miras de Araujo (VU)
Where is the robot?: Life as it could be.

2020-22 Maryam Masoud Khamis (RUN)
Understanding complex systems implementation through a modeling approach: the case of e-government in Zanzibar.

2020-23 Rianne Conijn (UT)
The Keys to Writing: A writing analytics approach to studying writing processes using keystroke logging.

2020-24 Lenin da Nobrega Medeiros (VUA/RUN)
How are you feeling, human? Towards emotionally supportive chatbots.

2020-25 Xin Du (TUE)
The Uncertainty in Exceptional Model Mining.

2020-26 Krzysztof Leszek Sadowski (UU)
GAMBIT: Genetic Algorithm for Model-Based mixed-Integer opTimization.

2020-27 Ekaterina Muravyeva (OUN)
Personal data and informed consent in an educational context.

2020-28 Bibeg Hang Limbu (OUN)
Multimodal interaction for deliberate practice: Training complex skills with augmented reality.

Open Universiteit