PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
SCHOOL OF ENGINEERING

PROCESS PERSPECTIVE IN MEDICAL EDUCATION

VÍCTOR ANDRÉS GÁLVEZ YANJARÍ

Thesis submitted to the Office of Graduate Studies
in partial fulfillment of the requirements for the degree of
Doctor in Engineering Sciences

Advisors:
MARCOS SEPÚLVEDA FERNÁNDEZ
JORGE MUÑOZ GAMA

Santiago de Chile, August 2022

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
SCHOOL OF ENGINEERING

PROCESS PERSPECTIVE IN MEDICAL EDUCATION

VÍCTOR ANDRÉS GÁLVEZ YANJARÍ

Members of the Committee:

MARCOS SEPÚLVEDA
JORGE MUÑOZ GAMA
VALERIA HERSKOVIC
EDUARDO KATTAN
VILMA MEJÍA
NIELS MARTIN
SERGIO MATURANA

Thesis submitted to the Office of Graduate Studies in partial fulfillment of the requirements for the Degree Doctor in Engineering Sciences

Santiago de Chile, August 2022

To my mum, dad and brother, who supported me all these years.

ACKNOWLEDGEMENTS

There are many ways to write this section. I will use it to thank for all the years I was at this university, which corresponds to twelve and a half years as a student (from 2010 at PENTA UC to the completion of this thesis).

First I thank the university. We know that no human construction is perfect, but if there is something close is the university. I thank to the university because it was the place where I met the most brilliant minds I have known so far, for helping me to develop my intellectual skills and acquire new knowledge in different areas, for opening my mind through the interaction with people from all social classes and parts of the world, and for helping me to satisfy the insatiable thirst for knowing more and more. I extend my thanks to all the professors, friends and classmates I met throughout these years.

Of course, I thank the university through the people I met there. Many will not be mentioned here, but I will name a few. With no doubt, the most important is Bernardita Román, who in the bus on the way to home encouraged me to bother my advisors to enter in the doctorate program. Special thanks to several of my closest friends (some of them are Hernán Scheel, Esaú Mendez, Mariajesús Espinosa, Rodrigo Rojas) for not doubting my abilities and encouraging me to enter the program. I also thank those people who told me "are you crazy?" "what are you doing!" "in my worst dreams I would spend four years more studying!" and similar things. I thank them because they made me feel that I was entering something really important, and therefore I made the decision with more reflection and seriousness.

Thanks go to Bernardo Pagnoncelli, probably the first person who made me think about the idea of doing a PhD. To David Acuña, a friend who was a couple of years ahead of me in graduate studies and was always available to advise me in evaluating whether doing a PhD was a good option. A special thanks also goes to Eric Rojas, who guided me

in my early undergraduate research and gave me the freedom to show that I was capable of doing research.

During these last four years I also met other colleagues who became friends, such as Daniela Silva from chemical engineering, Ana Maria Palomino from psychology, and probably the person with whom I shared most of the time in the pre-pandemic times in the office, Camilo Ruiz-Tagle. Special thanks to Camilo for his words of experience, professionally but also personally speaking. Thanks Camilo for the lessons and conversations about society, politics, music, cinema and so on, that helped me to discover new passions for life. Honorable mention to Juan Pablo Salazar and Gabriela Cajamarca, for the exchange of opinions and reciprocal help throughout these years.

"An atmosphere charged with enthusiasm and scientific efforts is the radical assumption for the existence of the university" [1] said a philosopher some time ago. It is exactly the spirit I found in every work meeting with my counterpart and fellow doctoral student René de la Fuente, and my supervisors Jorge Munoz-Gama and Marcos Sepúlveda. I really enjoyed all this time, because of the enthusiasm put by these three people in the research and the rigor required to run it. I have no words to thank these three people for the pleasant company these years, their total availability to help me in the thesis and for making this experience so pleasant.

Finally, I thank the State of Chile for financing my university education, particularly this academic degree. And finally, but clearly the most important, I thank my family, particularly my mum, dad and brother for their constant support and accompaniment in every decision I make.

[1]Citation from the book "Mission of the university", by José Ortega y Gasset. Own translation from the Spanish version of the quote in the book "Idea and defense of the university", by Jorge Millas (pg. 120, Ediciones Universidad Diego Portales).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Surgeons receive training in many skills. One of them is procedural skills, which enable physicians to perform surgical procedures. An important part of procedural skills is the sequence of steps that describes the order in which the actions involved in a surgical procedure should be performed. However, this aspect is often not reinforced or emphasised in procedural skills training, and typical instruments for assessing the learning of these competencies do not consider the sequence of steps.

Process Mining is a discipline coming from business process management, whose algorithms allow performing different tasks to support the processes that are executed in organisations from the data they leave in information systems. These algorithms help to discover a process model in order to know the sequence of steps in which a process is being executed, and to compare the executions of the process with the model that describes the ideal execution.

It has been proposed in the literature that a surgical procedure can be understood as a process, which enables the analysis of surgical procedures with process mining algorithms. Some studies have been conducted using process mining to understand the learning of surgical procedures, and it has been found that there is a high variability in the sequence of steps performed by residents during their training. However, it is not clear how the sequence of steps should be incorporated as a learning objective during resident training. Consequently, the learning of this aspect cannot be measured, and teaching it is difficult due to the absence of proven tools to do so.

This thesis seeks to provide tools to facilitate the teaching of procedural skills by instructors, helping them to incorporate the sequence of steps as a learning objective. The Process-Oriented Medical Education (POME) approach is proposed, which is composed of POME artifacts developed in the intersection of medicine, education and engineering.

This approach enables the teaching of the sequence of steps with solutions that consider the process perspective of procedural skills, which are called POME artifacts. Central venous catheter installation and percutaneous tracheostomy are used as case studies.

The thesis shows five contributions:

The first contribution is the identification of studies that explicitly reported the incorporation of the sequence of steps into procedural skills teaching and assessment strategies, through a systematic review of the literature. The results show the lack of research on this aspect in the medical education literature. The use of videos and the existence of non-standardised instruments to assess the learning of the sequence of steps were the strategies found.

The second contribution is a method to develop POME artifacts. This method consists of the creation of a process model that shows the consensus of the expert community about the procedure with respect to the sequence of steps in which it should be executed, the capture of data through the tagging of videos to obtain the sequence of steps performed by each recorded physician, and finally the development of an artifact to support a specific task to teach a surgical procedure using process mining algorithms.

The third contribution is a POME instrument to identify deficiencies in the learning of the sequence of steps. This instrument provides to instructors information on the deficiencies committed after a training session, which can be useful to make changes in the teaching strategy and identify parts that need to be practiced. The activities of central venous catheter installation were classified, and based on this classification the POME instrument was generated. Both developments were validated with experts (physicians with experience in instructing and performing this procedure), who found it useful for their training tasks.

The fourth contribution is a POME learning curve of the percutaneous tracheostomy's sequence of steps. With this artifact, instructors can know the residents' performance across the whole training course and have a sense of the learning progress of the course as a group. A similarity metric was used to compare residents' executions of the procedure with the process model of the ideal sequence of steps, for each resident and for each of the training course sessions. It was found that the curve reaches a plateau at the fifth session (there are seven in total). In addition, when analysing each stage of the procedure separately, it was found that there are stages that the residents learnt well and others that need to be reinforced.

The fifth contribution corresponds to POME metrics to determine the performance of the residents throughout the course at a low-level (in a more detailed way). Using data from the percutaneous tracheostomy course, the number of omissions, deviations and repetitions of activities were determined, with a detailed analysis by stage and by activity. These metrics were also validated by confirming the expected correlations with classical medical education metrics (time and OSATS), obtaining statistically significant values.

After developing POME artifacts to frame the POME approach, three statements were concluded. First, the POME approach provides useful artifacts to teach the sequence of steps. Second, developing POME artifacts requires process knowledge and medical sense. Third, residents do not perform the procedural skills used as case studies as they are supposed to, even at the training completion. Future work is needed to build and validate POME artifacts for training, assessment and feedback tasks, as well as to demonstrate its impact in learning.

Keywords: procedural skills; control-flow; process mining; medical education.

RESUMEN

Los cirujanos reciben entrenamiento en muchas habilidades. Una de ellas son las habilidades procedurales, las que permiten a los médicos realizar procedimientos quirúrgicos. Una parte importante de las competencias procedurales es la secuencia de pasos que describe el orden en que deben realizarse las acciones que involucran un procedimiento quirúrgico. Sin embargo, este aspecto suele no estar reforzado ni enfatizado en el entrenamiento de habilidades procedurales, y los instrumentos típicos para evaluar el aprendizaje de estas competencias no consideran la secuencia de pasos.

Minería de Procesos es una disciplina proveniente de la gestión de procesos de negocios, cuyos algoritmos permiten realizar distintas tareas para apoyar los procesos que se ejecutan en las organizaciones a partir de los datos que estos dejan en los sistemas de información. Estos algoritmos permiten descubrir un modelo de proceso para así conocer la secuencia de pasos en que se está ejecutando, así como también comparar las ejecuciones del proceso con el modelo que describe la ejecución ideal del proceso.

En la literatura se ha propuesto que un procedimiento quirúrgico puede entenderse como un proceso, lo que habilita el análisis de procedimientos quirúrgicos con algoritmos de minería de procesos. Se han realizado algunos estudios aplicando minería de procesos para entender el aprendizaje de procedimientos quirúrgicos, y se ha visto que existe una alta variabilidad en la secuencia de pasos que realizan los residentes durante su entrenamiento. Sin embargo, no está claro de qué manera debe incorporarse la secuencia de pasos como un objetivo a aprender durante el entrenamiento de residentes, por lo que no puede medirse su aprendizaje y se dificulta su enseñanza al no existir herramientas probadas para ello.

Esta tesis busca aportar con herramientas para facilitar la tarea educativa de los instructores de habilidades procedurales, ayudándoles a incorporar como un objetivo de

aprendizaje la secuencia de pasos. Se propone el enfoque POME (del inglés Process-Oriented Medical Education), el que está formado de artefactos POME desarrollados en la intersección entre medicina, educación e ingeniería. Este enfoque permite la enseñanza de la secuencia de pasos con soluciones que consideran la perspectiva de proceso de las habilidades procedurales, las que llamamos artefactos POME. Se usaron la instalación del catéter venoso central y la traqueostomía percutánea como casos de estudio.

La tesis muestra cinco contribuciones:

La primera contribución es la identificación de estudios que reportan explícitamente la incorporación de la secuencia de pasos en estrategias de enseñanza y evaluación de habilidades procedurales, a través de una revisión sistemática de la literatura. Los resultados muestran la necesidad de desarrollar artefactos POME para apoyar a los instructores, pues solo se encontraron nueve artículos. El uso de videos y la existencia de instrumentos no estandarizados para evaluar el aprendizaje de la secuencia de pasos son las estrategias encontradas más comunes.

La segunda contribución es un método para el desarrollo de artefactos POME. Este método consiste en el desarrollo de un modelo de proceso que muestra el consenso de la comunidad experta sobre el procedimiento con respecto a la secuencia de pasos en que éste debe ejecutarse, la captura de datos a través del etiquetado de videos para obtener la secuencia de pasos realizada por cada médico grabado, y finalmente, el desarrollo del artefacto requerido para apoyar una tarea específica de la enseñanza de habilidades procedurales usando algoritmos de minería de procesos.

La tercera contribución es un instrumento POME para identificar deficiencias en el aprendizaje de la secuencia de pasos. El instrumento entrega a los instructores información sobre las deficiencias cometidas después de una sesión de entrenamiento, lo que puede ser útil para hacer cambios a la estrategia de enseñanza e identificar partes que necesitan ser reforzadas. Se clasificaron las actividades de la instalación del catéter venoso central, y basándose en esta clasificación se generó un instrumento para instructores. Ambos desarrollos se validaron con expertos (médicos con experiencia en la instrucción y ejecución de este procedimiento), quienes lo encontraron útil para sus tareas educativas.

La cuarta contribución es una curva de aprendizaje de la secuencia de pasos. Con este artifacto, los instructores pueden conocer el desempeño de los residentes a largo del curso completo y tener una idea del progreso del aprendizaje a nivel de curso. Se ocupó una métrica de similaridad para comparar las ejecuciones del procedimiento realizadas por los residentes con el modelo de proceso que propone la secuencia ideal de pasos, para cada residente y para cada una de las sesiones del curso de entrenamiento. Se encontró que la curva alcanza una meseta a la quinta sesión (son siete en total), y al analizar cada etapa del procedimiento por separado se encontró que hay etapas que los residentes aprendieron bien y otras que necesitan reforzar al finalizar el curso.

La quinta contribución corresponde a métricas POME para determinar el desempeño de sus residentes a lo largo del curso a bajo nivel (de manera más detallada). Ocupando datos provenientes del curso de traqueostomía percutánea, se determinó el número de omisiones, desviaciones y repeticiones de actividades, haciendo un análisis detallado por etapa y por actividad. También se validaron estas métricas confirmando las correlaciones esperadas con métricas clásicas de educación médica (tiempo y OSATS), obteniendo valores estadísticamente significativos.

Como conclusión, luego de desarrollar artefactos POME y dar forma al enfoque POME, se pueden concluir tres afirmaciones. Primero, el enfoque POME provee artefactos útiles para enseñar la secuencia de pasos. Segundo, desarrollar artefactos POME requiere del conocimiento de procesos, pero también sentido médico. Tercero, los residentes no realizan las habilidades procedurales usadas como caso de estudio de la manera en que deberían, incluso al finalizar el curso. Como trabajo futuro, se propone construir y validar artefactos POME para las tareas de entrenamiento, evaluación y feedback, así como también demostrar su impacto en el aprendizaje.

Palabras Claves: habilidades procedurales; orden de ejecución de actividades; minería de procesos; educación médica.

Chapter 1

Introduction

1. INTRODUCTION

Medical education, particularly surgical education, is moving from the apprenticeship model to the competency-based model. The first model is based on the observation by residents of an instructor, who is an expert physician, performing the surgical procedure either in a simulation-based environment or a work-based environment. Then, the residents perform the procedure imitating the instructor (Walter, 2006). The second model focuses on the competencies that residents need to learn and demonstrate its acquisition at the end of a course (Frank et al., 2010).

Professional competency is "the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served" (Epstein & Hundert, 2002). Usually, surgeons are assessed during their training on the following competencies: procedure preparation, infection control, communication with the patient, team working, patient safety and procedural competency (McKinley et al., 2008). Therefore, procedural skills are a competency that every surgeon needs to have to perform surgical procedures successfully.

The process of teaching surgical skills has been described by Aydin et al. (2017) (see Figure 1.1). It is a cyclic process composed by five stages. First, instructors teach a competency (in this thesis, a procedural skill) to the residents (training stage). Second, instructors assess the residents' learning of the competency (assessment stage). Third, instructors give feedback to the residents about their learning (feedback stage). Fourth, instructors detect gaps or mistakes in residents' learning (deficiencies identification stage). Fifth, instructors analyze residents' performance along the course to make changes in their teaching strategy to enhance the training (performance stage).

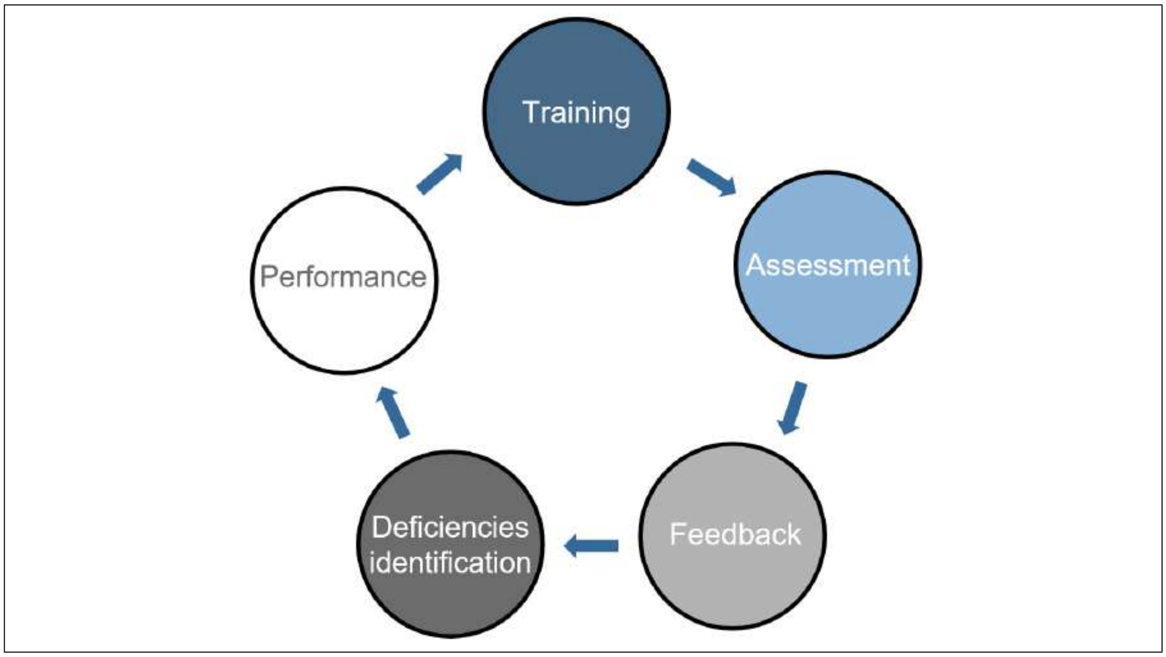


Figure 1.1. Procedural skill teaching cycle (adapted from Aydin et al. (2017)).

The learning of procedural skills requires the effort of cognitive, emotive and psychomotor capacities of residents (Bloom & Krathwohl, 1956). Psychomotor capacities are linked to the residents' expertise, and therefore to their clinical reasoning (which involves dimensions such as reflexivity, the ability to memorize and make inferences) (Higgs et al., 2008). Also, this capacity refer to the ability to perform movements (Oermann, 1990), and in surgery these movements need to be performed in a specific sequence of steps.

Shuell (1986) stated that learning is "an enduring change in behavior, or in the capacity to behave in a given fashion, which results from practice or other forms of experience". Schunk (2012) explained this statement decomposing it in three criteria: learning (i) involves change, which is seen in residents performing a procedure successfully in a different way (learning is inferential); (ii) endures over time, related to not forgetting how to perform a procedure in the long run (iii) occurs through experience and practice (excluding maturational changes in children).

The learning progress of the residents has been described as a spectrum in which they can improve their performance and reach a certain level of competence through training (ten Cate et al., 2010). According to Khan and Ramachandran (2012), the spectrum has seven levels:

1. Incompetent: the resident is unable to perform the procedure.
2. Novice: the resident considers each step isolated and it is difficult for him or her to deal with complexity/unexpected situations.
3. Advanced beginner: the resident considers the procedure as a sequence of steps and is able to deal partially with complex tasks.
4. Competent: the resident performs the procedure based on guidelines and previous experience. Also, the resident is able to deal with complexity.
5. Proficient: the surgeon performs the procedure mainly based on experience, following standards routinely and is able to see beyond the procedure.
6. Expert: the surgeon performs the procedure based on experience and intuition, in complex situations uses both analytical and intuitive systems.
7. Master: the surgeon performs the procedure easily in common situations, deals intuitively with complexity and establish new standards.

Hence, there is a need for instructors to teach the sequence of steps to novices, with the aim of helping them to move to the "advanced beginner" level of competence. The sequence of steps is the control-flow aspect of procedural skills, which pinpoint that surgical procedures can be treated as a collection of steps with a specific order (Lalys and Jannin, 2014; Neumuth, 2017). This statement makes the analysis of residents' performances of surgical procedures with process mining algorithms possible.

Process mining is a discipline composed of algorithms to understand processes using the data generated and stored in information systems that commonly support processes (van der Aalst, 2016). An overview is provided in Figure 1.2. These algorithms are typically classified into three categories: (a) Discovery, a representative process model is obtained from process executions; (b) Conformance Checking, to reveal discrepancies between a process model and process executions; (c) Enhancement, to extend, redesign or make improvements on a process model based on process executions. Process executions are stored as data in information systems while processes are running, which are transformed into "event logs". Event logs should have at least three types of data for process mining algorithms: (i) case id, the process execution identifier; (ii) activity, each action or step that the process has; (iii) timestamp, time the activity took place. Process mining has been successfully used in different healthcare specialties and provides opportunities to support instructors of procedural skills training throughout the procedural skill teaching cycle.

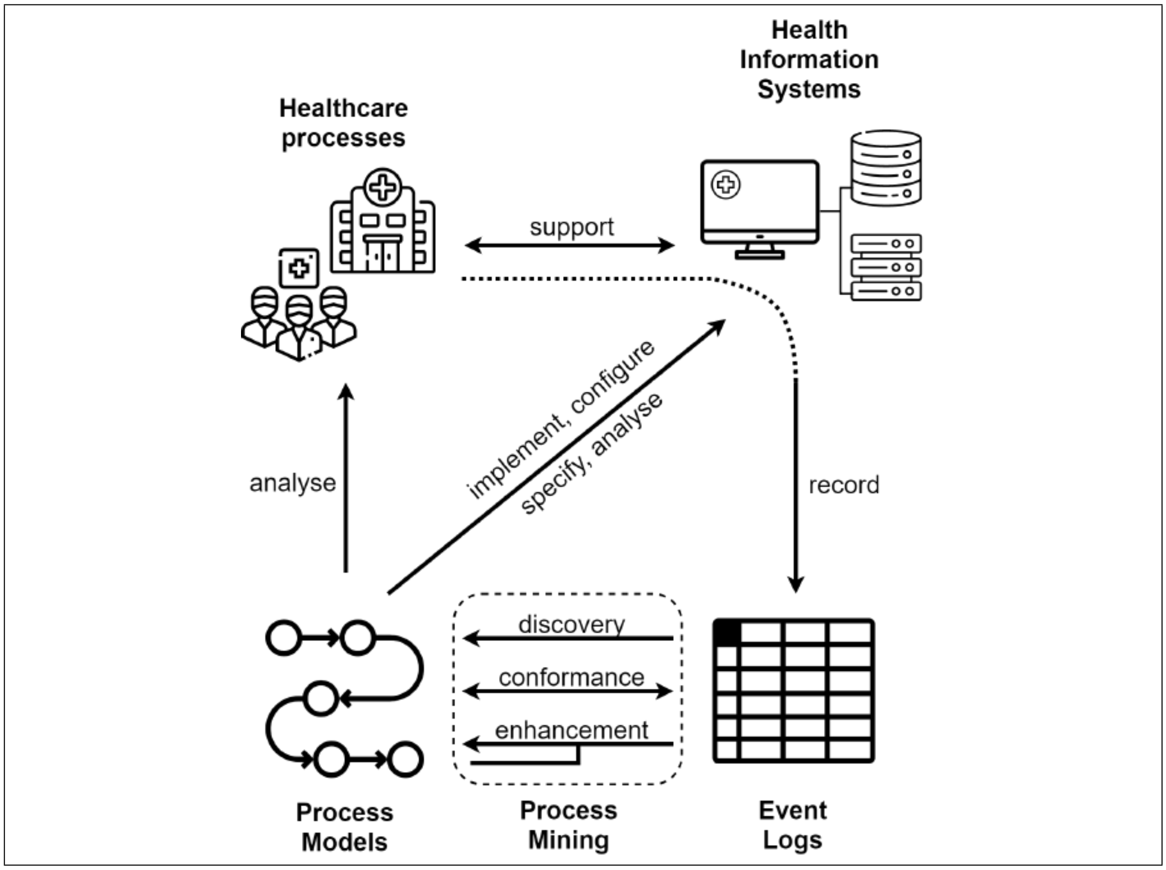


Figure 1.2. Overview of process mining in healthcare (extracted from Munoz-Gama et al. (2022)).

Instructors need to specify learning objectives (Berman, 2015) and information to run the procedural skill teaching cycle (Figure 1.1) with residents at the novice competence level.. Also, instructors are interested in detecting learning difficulties, getting a better understanding of the residents' learning process and discover learning flows (Bogarín et al., 2018). This thesis proposes a POME (Process-Oriented Medical Education) approach to support instructors in the tasks mentioned when teaching procedural skills to novices. The POME approach is a didactic method (Rivilla et al., 2009) born at the intersection between medical education, procedural skills and process mining to provide a control-flow perspective to teach procedural skills to novices in medical education, based on insights gained through process mining analyses. As surgical procedures can be seen as processes (Lalys and Jannin, 2014; Neumuth, 2017), the POME approach allows instructors to teach a procedural skill as if it were a process, helping residents to improve their performance moving from the novice to the advanced beginner competence level. It is known that the control-flow perspective of business processes can be analyzed and monitored (Dumas et al., 2018). Therefore, the POME approach helps instructors to analyze the control-flow performance of novices learning a surgical procedure, make changes in their teaching strategies to improve residents' performance and monitor the compliance of the residents' performance with the expected sequence of steps of the surgical procedure being taught.

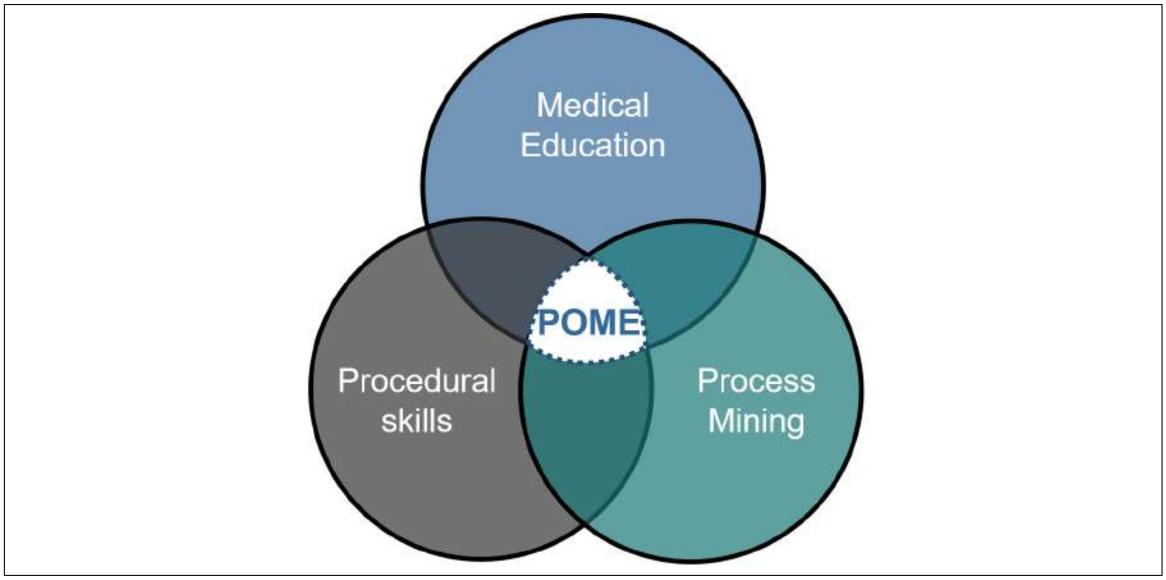


Figure 1.3. The POME (Process-Oriented Medical Education) approach emerges in the intersection between process mining and medical education to teach procedural skills to novices.

The POME approach is composed of a variety of tools to help instructors in their teaching tasks, which are called POME artifacts. Following Wieringa (2014), an artifact is a solution for a problem in a specific context. Hence, a POME artifact is a process-oriented solution (i.e. a solution that considers the control-flow perspective) to each stage of the procedural skill teaching cycle, in the context of procedural skills training.

A method to generate POME artifacts is described in this thesis, and using it a POME instrument, a POME learning curve and POME metrics were generated to support instructors. We understand as a method a set of steps and procedures to obtain a specific result, which in this thesis correspond to POME artifacts; a POME instrument is any artifact showing the control-flow performance of residents by any means (e.g. a report showing process-oriented metrics after a training session); a POME learning curve seeks to describe the learning progress of a procedural skill across a whole training course; and POME metrics are measurements useful to understand the control-flow performance of residents performing a procedural skill. The POME artifacts can be incorporated in teaching strategies, which are any action performed to teach a procedural skill throughout the procedural skill teaching cycle (Cook et al., 2013).

The POME approach is an opportunity to support instructors throughout the procedural skill teaching cycle. Challenges, objectives and contributions to shape this approach are presented below.

1.1. Challenges

As previously mentioned, instructors are in need of standard information to run the procedural skill teaching cycle and establish specific learning objectives (Berman, 2015) . Also, the sequence of steps is an inherent aspect of surgical procedures (Oermann, 1990) and, as such, is commonly considered an easy-to-learn feature. However, previous research have shown differences between novices and experts on the sequence of steps they perform in a simulation context (de la Fuente et al., 2020). Besides, it has been seen that residents don't perform the sequence of steps as expected after a training session (Lira et al., 2019).

Therefore, to support instructors throughout the procedural skill teaching cycle, this thesis addresses the following challenges:

[Challenge 1] Instructors need to understand how the control-flow perspective can be explicitly incorporated in procedural skills training when teaching novices.

Checklists and global rating scales are instruments commonly used in procedural skills training (Epstein, 2007). However, these instruments omit aspects of procedural skills that surgeons should know for their daily practice (McKinley et al., 2008). Therefore, determining whether the control-flow perspective is one of those aspects and how it can be incorporated is crucial to help instructors include this aspect in training courses.

[Challenge 2] Instructors need to identify the students' *deficiencies* about control-flow at the group level and at the student level, to improve future training sessions.

It is difficult for instructors to detect residents' mistakes, especially when it comes to control-flow mistakes (Sullivan et al., 2014). Hence, there is a need for developing tools to help instructors understand control-flow mistakes while residents are in training.

[Challenge 3] Instructors need to determine the students' *performance* about control-flow throughout the training sessions to determine procedural skill evolution.

The learning trajectory of a procedural skill is relevant for instructors. Such information is helpful to redesign sessions and define the residents' strengths and flaws along the course (Pusic et al., 2015). Despite this, there is no study showing the control-flow learning evolution of any surgical procedure.

1.2. Research Question, Hypothesis and Objectives

The following research question, hypothesis, general objective and specific objectives are formulated to address the challenges proposed in section 1.1:

1.2.1. Research Question

To address the challenges proposed in section 1.1, the following research question was formulated: How can the POME approach support instructors on the understanding of control-flow *deficiencies* and *performance* of the residents at a novice level in a procedural skills training?

1.2.2. Hypothesis

The hypothesis that drives the work done in this thesis is that using POME artifacts allow instructors to understand control-flow *deficiencies* and *performance* of the residents at a novice level in a procedural skills training.

1.2.3. General Objective

The objective defined to answer the research question is to develop POME artifacts that allow instructors to understand control-flow *deficiencies* and *performance* of the residents at a novice level in a procedural skills training.

1.2.4. Specific Objectives

The general objective was divided in three specific objectives:

[Objective 1] Identify existing artifacts in the literature to explicitly teach and assess the control-flow in procedural skills training, and define the most suitable POME artifacts to do so.

[Objective 2] Design a POME artifact to help instructors understand residents' deficiencies about control-flow, at the group and student level.

[Objective 3] Design a POME artifact to study the residents' performance in terms of control-flow throughout the procedural skills training, to visualize the learning of the procedural skill.

1.3. Contributions

This thesis contributes to shape the POME approach by generating POME artifacts to support instructors through process mining algorithms across the procedural skill teaching cycle. All these artifacts provide a concrete, structured and tangible way to teach an inherent aspect of procedural skills (i.e., the sequence of steps) that has not been considered as a first-class citizen in medical education. Also, the contributions provide insights and tools to give the sequence of steps relevance on training courses as a specific learning objective. This aim is accomplished through the quantification of sequence errors made by novices during the training sessions.

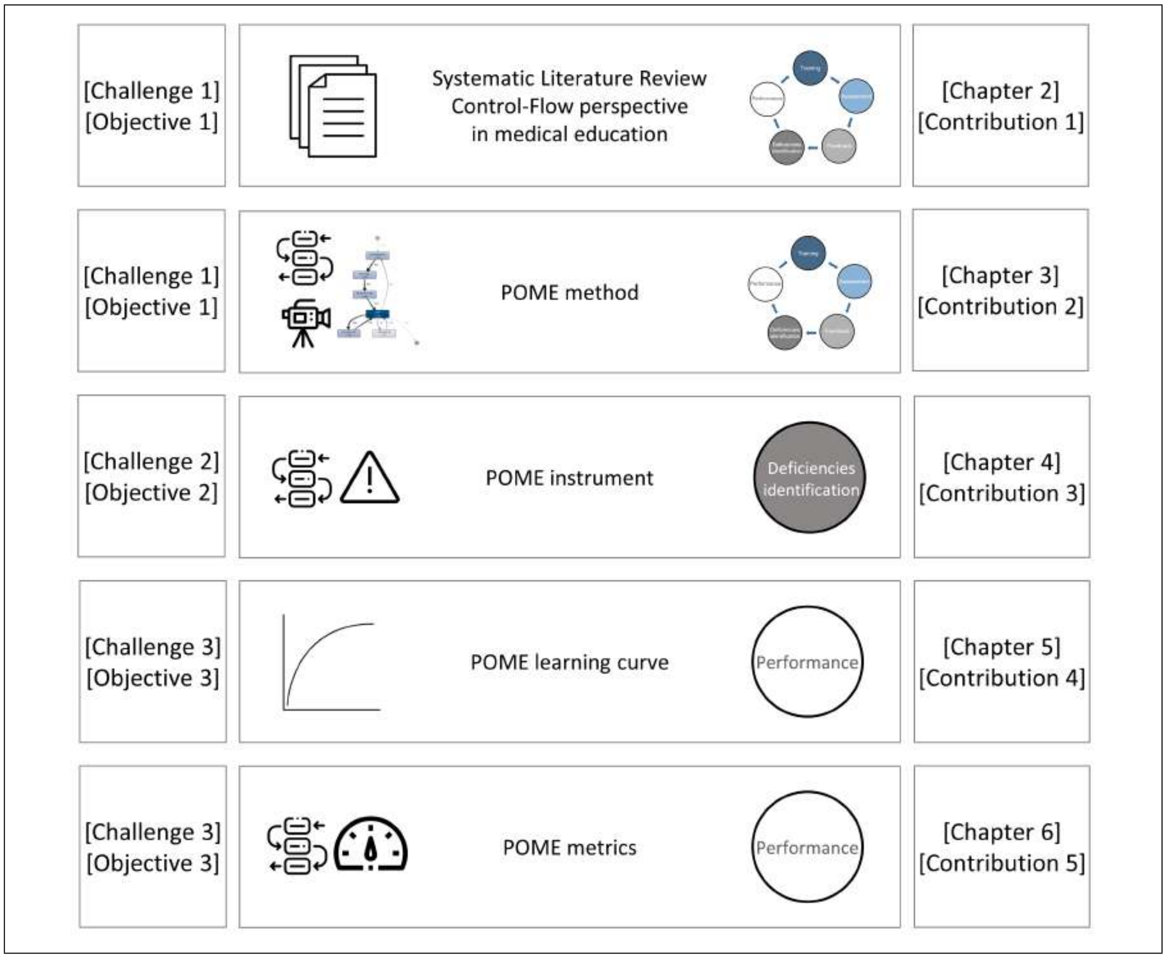


Figure 1.4. Thesis overview. Each row of the figure corresponds to a contribution. The first column corresponds to the challenge tackled and the specific objective completed by each contribution. The second column summarizes each contribution with a figure and one sentence, and also shows the stage of the procedural skill teaching cycle that the contribution seeks to support. Finally, the third column shows the chapter and the corresponding contribution number.

Table 1.1. Summary of contributions.

Contribution	Objective	Challenge	Chapter	Publications
[Contribution 1] A literature study of the existing artifacts that consider the control-flow perspective.	[Objective 1]	[Challenge 1]	[Chapter 2]	[Systematic Literature Review]
[Contribution 2] A method to develop POME artifacts.	[Objective 1]	[Challenge 1]	[Chapter 3]	[POME method]
[Contribution 3] A POME report-type instrument to identify control-flow deficiencies.	[Objective 2]	[Challenge 2]	[Chapter 4]	[Process-Oriented Instrument (POI)] - [POI Conference paper]
[Contribution 4] A POME learning curve comparing the performance of residents versus an ideal execution of the procedure.	[Objective 3]	[Challenge 3]	[Chapter 5]	[Learning Curve]
[Contribution 5] POME metrics to determine residents' control-flow performance.	[Objective 3]	[Challenge 3]	[Chapter 6]	[Learning Curve Top-Down Analysis]

[Contribution 1] A literature study of the existing artifacts that consider the control-flow perspective.

It is needed a standardization on how to teach the sequence of steps to novices, which involves exposing them to different ways to perform a surgical procedure and to different contexts that they will encounter in real-life scenarios (Apramian et al., 2015, 2016). Hence, a systematic literature review of studies that considered an explicit inclusion of the control-flow perspective of surgical procedures in teaching and assessment of procedural skills was conducted, following the PRISMA guidelines (Moher et al., 2009). The results show that there is still a need for research on this aspect, since we found few articles and there were high variability in the tools and methods instructors use to teach and assess the sequence of steps. This contribution is described in Chapter 2 of this thesis.

[Contribution 2] A method to develop POME artifacts.

Developing POME artifacts has no precedents in the literature. However, how to describe processes (Dumas et al., 2018), data collection in medical education (Green et al., 2019) and artifact development (Wieringa, 2014) are extensively described. Therefore, mixing these three areas contributes with a method to create tangible artifacts to quantify sequence of steps' errors of novices in medical education. Creating POME artifacts require a specific setting to generate process data and then analyze it. The method presented in this thesis consists of establishing a consensus on the ideal process execution, obtaining the control-flow data from videos and producing POME artifacts to support instructors throughout the procedural skills teaching cycle. This contribution is described in Chapter 3 of this thesis.

[Contribution 3] A POME report-type instrument to identify control-flow deficiencies.

With the POME method it is possible to design artifacts to quantify the sequence errors that novices make during a training course. There is no structured nor quantitative way in the literature to identify control-flow deficiencies (common techniques to do so are the instructors' observation of residents performing the procedure and immediate feedback (Lammers, 2008)). Thus, an instrument to support instructors in detecting control-flow deficiencies was developed, using data from a central venous access guided by ultrasonography training course of the anesthesiology and intensive medicine residency program (ethical approval (ID 16-194) by the Pontificia Universidad Católica de Chile ethics research committee). Also, a validation with instructors who teach this surgical procedure was conducted. This contribution is described in Chapter 4.

[Contribution 4] A POME learning curve comparing the performance of residents versus an ideal execution of the procedure.

Learning curves are helpful to describe skill acquisition across a training course (Pusic et al., 2015). However, these curves are typically built on real clinical settings and using classic metrics such as performance time duration, checklists and global rating scales (Massick et al., 2000; Petiot et al., 2017; Kattan et al., 2020). This thesis presents a learning curve built on a metric that measures the sequence of steps' performance that novices performed across a simulation-based training course. A learning curve was built comparing residents' performance in the bronchoscopy-guided percutaneous dilatational tracheostomy installation training course of the intensive medicine residency program (ethical approval (ID 180704005) by the Pontificia Universidad Católica de Chile ethics research committee) with the ideal surgical process model (de la Fuente et al., 2021). This analysis helps to determine the control-flow performance for the whole surgical procedure and for each stage. This contribution is described in Chapter 5.

[Contribution 5] POME metrics to determine residents' control-flow performance.

The learning curve is a clear contribution to the field. However, it was designed to provide global insights and at high level. Since instructors are in need of detailed information of procedural performance (Bogarín et al., 2018), POME metrics were developed to quantify sequence errors and obtain information about the sequence mistakes residents commit across a training course. A top-down analysis of the bronchoscopy-guided percutaneous dilatational tracheostomy installation training course of the intensive medicine residency program (ethical approval (ID 180704005) by the Pontificia Universidad Católica de Chile ethics research committee) was conducted to discover the control-flow performance through process-oriented metrics, to provide detailed information about the learning of the control-flow perspective. Also, process-oriented metrics' behavior was compared with global rating scales results. This contribution is described in Chapter 6.

1.4. Impact

The impact of this thesis is presented below, classified in research and educational impact.

1.4.1. Research impact

The work presented in this thesis was made available to the research community through journal papers, a book chapter, a conference presentation and a research visit, which are detailed below.

Journal Papers

[Process-Oriented Instrument (POI)] Galvez, V., de la Fuente, R., Meneses, C., Leiva, L., Fagalde, G., Herskovic, V., Fuentes, R., Munoz-Gama, J., & Sepúlveda, M. (2020). Process-Oriented Instrument and Taxonomy for Teaching Surgical Procedures in Medical Training: The Ultrasound-Guided Insertion of Central Venous Catheter. *International Journal of Environmental Research and Public Health*, *17*(11). https://doi.org/10.3390/ijerph17113849

[TDA] Martínez, J. J., Galvez-Yanjari, V., de la Fuente, R., Kychenthal, C., Kattan, E., Bravo, S., Munoz-Gama, J., & Sepúlveda, M. (2022). Process-oriented metrics to provide feedback and assess the performance of students who are learning surgical procedures: The Percutaneous Dilatational Tracheostomy case. *Medical Teacher*. https://doi.org/10.1080/0142159X.2022.2073209

Book Chapter

[POME method] Munoz-Gama, J., Galvez, V., Fuente, R. de la, Sepúlveda, M., & Fuentes, R. (2021). Interactive Process Mining for Medical Training. In C. Fernandez-Llatas (Ed.), *Interactive Process Mining in Healthcare* (pp. 233–242). Springer, Cham. https://doi.org/10.1007/978-3-030-53993-1

Conference Presentation

[POI Conference paper] Galvez, V., Meneses, C., Fagalde, G., Munoz-Gama, J., Sepúlveda, M., Fuentes, R., & de la Fuente, R. (2019). Understanding Undesired Procedural Behavior in Surgical Training: The Instructor Perspective. In Di Francescomarino, C., Dijkman, R., Zdun, U. (Eds), *Business Process Management Workshops. BPM 2019. Lecture Notes in Business Information Processing, vol 362*. Springer, Cham. https://doi.org/10.1007/978-3-030-37453-2_38

Under Review Papers

[Systematic Literature Review] Galvez-Yanjari, V., de la Fuente, R., Munoz-Gama, J., & Sepúlveda, M. (2021). Procedural Variability and the Sequence of Steps in Procedural Skills Training: A Systematic Literature Review. *Simulation in Healthcare*.

[Learning Curve] Galvez-Yanjari, V., Kattan, E., de la Fuente, R., Kychenthal, C., Munoz-Gama, J., & Sepúlveda, M. (2021). Percutaneous Dilatational Tracheostomy Training from Control-Flow Perspective: The parts do not equal the whole. *Canadian Journal of Anesthesia*.

Research visit

The author of this thesis visited and conducted research in the Business Process Management and Analytics lab at Utrecht University, The Netherlands, under the supervision of Hajo A. Reijers from January 2022 to May 2022. During the stay, the author of this thesis worked in collaboration with Jan Willem (Syrius Medical company) and Bart Vrouenraets, MD (OLVG West hospital, Amsterdam) to improve the breast conserving cancer surgery process. As a result, recommendations were generated to improve the patient comfort across the process through a process mining analysis.

1.4.2. Educational impact

- The POME approach highlights the relevance of an aspect typically disregarded in procedural skills training, the control-flow perspective. The research presented in this thesis is a direct input for instructors to teach this aspect and use it throughout the procedural skill teaching cycle.
- Instructors can use the knowledge provided by this thesis to make improvements to their training and decide whether there are some parts in need of additional training sessions.
- As procedural skills training is transitioning to a competency-based model, this thesis produces a variety of POME artifacts (a method, an instrument, a learning curve and metrics) to help instructors to teach a specific competency aspect, i.e., the control-flow perspective of procedural skills.
- The instruments typically used in medical education to assess competencies' learning are based on expert knowledge, i.e., the instructor is evaluating the resident based on perception, which can be biased. This thesis provides different POME artifacts to quantify the control-flow of surgical procedures and thus avoid bias.

1.5. Thesis structure

The thesis is composed of chapters describing the research contributions, which are available to the research community in different formats. Chapter 1 is the thesis introduction; chapters 2, 4, 5 and 6 were submitted to WoS journals (chapter 4 and 6 are published, 2 and 5 are under review), chapter 3 was published as a book chapter, and chapter 7 describes the conclusions and future work. The chapters and a summary of their content are described below.

Chapter 2: Procedural Variability and the Sequence of Steps in Procedural Skills Training: A Systematic Literature Review (Galvez-Yanjari, V., de la Fuente, R., Munoz-Gama, J., & Sepúlveda, M. (2021). ***Simulation in Healthcare*****, under review.).**

This chapter presents the systematic literature review of studies reporting strategies to teach and assess the control-flow perspective of surgical procedures. Also, it provides recommendations to include this perspective in procedural skills training. After searching Embase, PubMed, Web of Science, Google Scholar and CINAHL (Cumulative Index to Nursing and Allied Health Literature) databases, 4326 articles were collected. 9 met the inclusion criteria, which were assessed using the MERSQI (Medical Education Research Study Quality Instrument) scale. We found videos demonstrating the procedure and giving immediate feedback about the control-flow perspective as teaching strategies. Adherence to a predefined order and omission of steps were found as outcomes to assess the learning of the control-flow perspective. We concluded that the control-flow perspective is rarely reported in procedural skills teaching and assessment studies, and suggested the inclusion of process models depicting surgical procedures to teach this perspective.

Chapter 3: Interactive Process Mining for Medical Training (Munoz-Gama, J., Galvez, V., Fuente, R. de la, Sepúlveda, M., & Fuentes, R. (2021). In C. Fernandez-Llatas (Ed.), ***Interactive Process Mining in Healthcare*** **(pp. 233–242). Springer, Cham (link to paper).**

This chapter describes a method to incorporate the control-flow perspective in procedural skills training, based on process models and process mining. The method has three stages: (1) model stage, where a process model is designed after a literature review where the steps a surgical procedure should have are identified, to then make changes to the initial process model based on experts' opinion according to a Delphi panel; (2) data stage, which consists of recording surgeons performing the surgical procedure, either in a simulation or real environment, to later tag the video with the steps executed by the surgeon to obtain event logs; (3) analysis stage, where process mining analyses are conducted depending on the stage of the procedural skill teaching cycle at which instructors need support.

Chapter 4: Process-Oriented Instrument and Taxonomy for Teaching Surgical Procedures in Medical Training: The Ultrasound-Guided Insertion of Central Venous Catheter (Galvez, V., de la Fuente, R., Meneses, C., Leiva, L., Fagalde, G., Herskovic, V., Fuentes, R., Munoz-Gama, J., & Sepúlveda, M. (2020). *International Journal of Environmental Research and Public Health*, *17*(11) (link to paper)).

This chapter exposes the generation of a taxonomy to classify surgical steps according to the high-level tasks of surgical procedures (prepare, identify, act, verify). Also, it exposes the development of a POME report-type instrument to support instructors in detecting deficiencies, which is based on the steps' classification proposed by the taxonomy. Both developments were illustrated using data of ten residents in a UGIJCVC (Ultrasound-Guided Internal Jugular Central Venous Catheter) simulation-based training course, and validated with three instructors who have 5.7 years of experience teaching the UGIJCVC procedure through a test to assess their understanding of the information the instrument contains (instrument interpretability test), a usability test and instructors' opinions. The results showed that instructors mostly understood the information presented in the instrument, the instrument had an acceptable level of usability, and instructors found that both the taxonomy and the instrument are helpful.

Chapter 5: Percutaneous Dilatational Tracheostomy Training from Control-Flow Perspective: The parts do not equal the whole (Galvez-Yanjari, V., Kattan, E., de la Fuente, R., Kychenthal, C., Munoz-Gama, J., & Sepúlveda, M. (2022). *Canadian Journal of Anesthesia*, under review.)

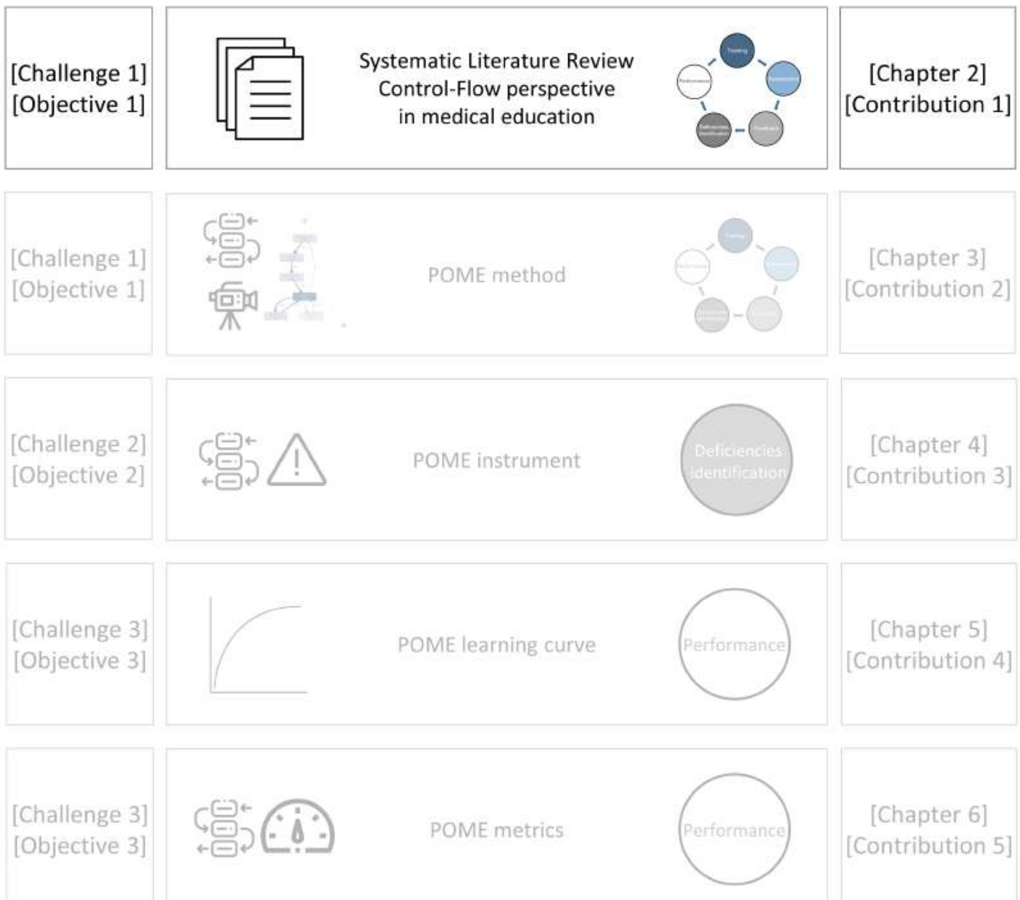
This chapter proposes the construction of a learning curve to obtain insights about the learning of the control-flow perspective by eight anesthesia and intensive care medicine residents learning the PDT (Percutaneous Dilatational Tracheostomy) procedure in a seven-sessions simulation-based training course. Normalized Levenshtein Distance (NLD) was used to compare the sequence of steps followed by residents in each session versus the process model representing the surgical procedure. The results showed that residents improve their learning as the course progresses through the sessions, and NLD variability decreases as the course progresses as well. However, the analysis by stage indicates that at the end of the training there were still parts of the procedure that residents need to rehearse (tracheal puncture and tracheal dilatation stages).

Chapter 6: Process-oriented metrics to provide feedback and assess the performance of students who are learning surgical procedures: The Percutaneous Dilatational Tracheostomy case (Martínez, J. J., Galvez-Yanjari, V., de la Fuente, R., Kychenthal, C., Kattan, E., Bravo, S., Munoz-Gama, J., & Sepúlveda, M. (2022) *Medical Teacher* (link to paper)).

This chapter details process-oriented metrics to get information about residents' performance through a top-down analysis, using data of eight residents in a PDT seven-sessions simulation-based training course. Three process-oriented metrics to discover procedural mistakes were calculated (omission, repetition, deviation) and compared with global rating scores and execution time. The results showed that procedural mistakes decreased as the course progressed when the procedure was analyzed as a whole. However, the analysis by stage showed that only one stage (preparation) had no errors at the end of the training. Also, after focusing on stages with no clear tendency (tracheal puncture and tracheal dilatation stages), it was possible to analyze the residents' performance for each activity along the course, identifying the most and least involved activities in control-flow mistakes. Additionally, the three process-oriented metrics were compared to classic metrics (total time duration and global rating scores) with the pearson's coefficient index, obtaining statistical significant values. The comparison was conducted to evaluate whether the process-oriented metrics behave as expected. In particular, there was a positive correlation with total time duration, and a negative correlation with global rating scores.

Chapter 2

Procedural Variability and the Sequence of Steps in Procedural Skills Training: A Systematic Literature Review



2. PROCEDURAL VARIABILITY AND THE SEQUENCE OF STEPS IN PROCEDURAL SKILLS TRAINING: A SYSTEMATIC LITERATURE REVIEW

2.1. Introduction

Surgeons receive training on a wide variety of skills. One of them are procedural skills, which are composed of affective, cognitive and psychomotor domains (Bloom & Krathwohl, 1956). These domains are complex to develop because they are interdependent on each other (Menix, 1996; Shaker, 2018). In particular, the psychomotor domain includes the ability to perform skills that involve movements (Oermann, 1990), and Dave's taxonomy establishes that it is developed by imitating the skill to perform the movements in a logical sequence naturally (da Costa et al., 2018).

Procedural variability refers to differences on how surgeons perform procedures (Apramian et al., 2015, 2016). These variations occur due to different surgeons' backgrounds and individual preferences (Golden et al., 2020). It is present when surgeons embrace challenging situations in the operating room, and even when procedures are practiced on simulation-based contexts, where external factors such as resources and patient variability do not exist (Golden et al., 2020). Particularly, it has been shown that variability in the sequence of steps, which is related to the psychomotor domain of procedural skills, is the rule and not the exception (Neumuth, 2009; Rojas, Munoz-Gama, Sepúlveda, and Capurro, 2016; Neumuth, 2017, de la Fuente et al., 2020). Furthermore, instructors believe that dealing with procedural variability is a relevant aspect that residents need to learn (Apramian et al., 2016), but it is difficult for residents to recognize what is a principle (i.e, a mandatory step) and what is an individual preference (i.e., options to perform a step) (Apramian et al., 2015) in every different way to perform the procedure. Hence, the sequence of steps needs to be considered in training programs to improve residents' learning (Apramian et al., 2015; Légaré et al., 2015; Apramian et al., 2016).

Standardization of surgical procedures is key to make residents deal with procedural variability in training programs, which is positive to increase the surgical quality and the patient safety (Leotsakos et al., 2014; Fecso et al., 2017; Skjold-Ødegaard & Søreide, 2020). This process involves the consideration of the different paths a surgical procedure can have and decision making on what is relevant to include in the standard generated (Neumuth, 2017; de la Fuente et al., 2020). The sequence of steps' standardization allows to clearly establish the competence that residents must acquire at the end of the training, which helps to design novel strategies to be used in competency-based training programs (Frank et al., 2010). For this task, it is known that surgical procedures can be represented as a set of sequentially ordered steps (Lalys and Jannin, 2014; Neumuth, 2017). However, no systematic review had focused on describing studies that reported the explicit incorporation of a standard sequence of steps in training programs, particularly into teaching and assessment strategies of procedural skills training.

The objective of this research was two-fold. First, to conduct a systematic review of articles that reported strategies considering the teaching and assessment of a standard sequence of steps in procedural skills training. Second, to present recommendations to include the sequence of steps into procedural skills teaching and assessment training strategies.

2.2. Methods

2.2.1. Research questions

This review seeks to answer the following questions: (a) What strategies have been reported in instructional designs' studies to teach a standard sequence of steps in procedural skills training? Are these strategies effective? (b) What strategies, instruments and outcomes have been reported to assess the sequence of steps' learning in procedural skills training? Is there validity evidence for these instruments?

To answer these questions, we understood as a strategy any action performed to explicitly teach or assess the sequence of steps (e.g., showing in a printed flowchart the standard sequence or saying aloud the steps in order) (Cook et al., 2013). Also, we understood as an instrument any tool to assess the sequence of steps' learning.

2.2.2. Protocol

To conduct this systematic review, we created a protocol following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al., 2009).

2.2.3. Eligibility criteria

We included original and full-text articles reporting that participants were explicitly taught about a standard sequence of steps, or the assessment of learning considered whether participants' performance adhered to a standard sequence of steps. Particularly, we included articles reporting that the standard sequence of steps was taught using any strategy to make it explicit, or the assessment considered whether students omitted steps, repeated steps or performed steps in the wrong sequence. Also, we included studies where the participants were faculty or students at any level. We did not restrict articles' inclusion to any specific specialty, study design or training level (undergraduate, postgraduate or staff).

The eligibility criteria we used are the following:

- Population: doctors (instructors or students) who perform a procedure for training and/or assessment purposes.
- Intervention: any with the aim of training and/or assessment of procedural skills.
- Comparison: those who don't participate of the intervention (control v/s intervention group), in case of evaluation of interventions. If it is a design development, the current literature.
- Outcomes: formalization/representation of the procedure; performance analysis of the procedure; development, validation and/or evaluation of tools or instruments.
- Study design: the ones that allow to get the outcomes described before.

We excluded articles that considered teaching or assessing the learning of each step separately. We also excluded abstracts, conference papers, reviews, editorials, opinion letters, and articles written in a language other than English.

2.2.4. Search strategy

We searched EMBASE, PubMed, Web of Science, Google Scholar, and CINAHL databases to ensure good literature coverage (Bramer et al., 2017). We searched each database from the inception to September 30, 2019, and we used the following search strategy: ('procedural') AND ('skill' OR 'skills' OR 'competence' OR 'competency') AND ('training' OR 'teaching' OR 'instruction'

OR ‘assessment’). We selected the term ‘procedural’ because this is the aspect that contains the sequence of steps of surgical procedures. The term ’skills’ was associated with other synonyms that can help to catch papers related to education of procedural skills. Finally, the last part of the string helped to focus the search on the stages that we are interested in, teaching and assessment. We used broad search criteria to maximize the number of potential candidate articles. In the Google Scholar database, patents and citations were excluded.

2.2.5. Study selection

One author (V.G-Y.) screened titles and abstracts. After removing duplicates, the same author retrieved articles eligible for full-text reading. Then, two authors (V.G-Y. and R.d.l.F.) assessed the articles’ eligibility, resolving disagreements through discussion. Additionally, we conducted a backward search (Moher et al., 2009), i.e., we manually searched the references list of included articles for possible new eligible articles.

2.2.6. Data collection

One author (V.G.) collected the data by completing a form for each included article; then, the data was verified by a second author (R.d.l.F.). We classified the data extracted by specialty, procedure, instructional modality, study details (such as study design and participants), and aspects of the sequence of steps (whether the study reported that the sequence of steps was explicitly taught or assessed, strategies reported to explicitly teach or assess the sequence of steps, outcomes, results, validity of instruments to assess the sequence of steps). We collected and managed the data in Google Sheets, and we performed a qualitative synthesis to analyze the data. We did not perform meta-analysis due to the variety of study designs and outcomes found in the included articles.

2.2.7. Validity of assessment’s instruments

To determine the validity of the assessment instruments found in this review, we searched evidence using the practical guidelines proposed by Borgersen et al. (2018). For content validity, they proposed to consult people with experience in the procedure to design or make adjustments to an instrument. For response validity, they suggested using standardized instructions and blinded raters to minimize biases. For internal structure validity, they recommended determining the reliability of scores (i.e., the instrument provides the same results each time it is used to assess residents) through statistical methods. For relationship to other variables, they proposed to determine the correlation between assessment scores and other variables, such as experience or proficiency level. Finally, for consequential validity, they proposed to determine the consequences that the test had, for example, defining a pass-fail score.

2.2.8. Quality assessment

Two authors (V.G. and R.d.l.F.) assessed the quality of the included articles independently, and resolved disagreements through discussion. We used the Medical Education Research Study Quality Instrument (MERSQI) (Reed et al., 2007; Cook & Reed, 2015), a rating scale to assess the methodological quality of quantitative studies in medical education research. With this instrument, it was possible to rate the study design, sampling, type of data, validity, data analysis, and outcomes of each article (Reed et al., 2007; Cook & Reed, 2015). Also, MERSQI defines a maximum score of 3 points for each component, and rates each study with a maximum total score of 18 points.

2.3. Results

2.3.1. Study selection

Figure 2.1 shows the PRISMA flow diagram. In total, we identified 7175 articles by searching the databases. 4346 articles remained after removing duplicates.

As a result of a first titles and abstracts screening using the eligibility criteria mentioned earlier, we selected 233 articles for a second screening. We discarded 2287 titles and abstracts because the intervention was not for procedural skills training or assessment purposes, 1151 that had a topic different than medical education, 337 because the outcome was not the desired, 185 reviews, 74 that had participants different than doctors, 21 abstracts, 16 posters, 12 editorials, 11 not accessible, 7 books, 5 letters to the editor, 3 corrections to papers, 3 replies to papers, and 1 non-english-written paper.

After a second screening of the 233 articles filtered, we selected 17 of them for full-text assessment. We discarded 79 titles and abstracts because they talk about teaching and/or assessing each step isolated (i.e., not the sequence), 48 abstracts, 33 that talked about defining the procedural steps, 29 that had an intervention with purposes different to training or assessment, 8 with undesired outcomes, 7 posters, 5 that had a topic different than medical education, 5 not accessible articles, 1 review and 1 thesis.

After the full-text assessment of the 17 articles, we selected 8 articles for inclusion, and we excluded 9 articles: 2 because the assessment instrument does not allow evaluating whether residents made sequence errors, 2 because the teaching strategy focused on each step separately and not on the sequence of the steps, 2 where the authors determined the frequency of each procedural step (not the sequence), 2 that used a machine learning algorithm for tasks other than teaching and assessing the sequence of steps, and 1 because it was a conference paper. We included 1 article after the manual search in the references list of the articles fully read.

2.3.2. Characteristics of included articles

A detailed overview of the included articles is presented in Table 2.1. The included articles considered procedures from emergency medicine (Chapman et al., 1994; Lammers et al., 2008), pediatry (Lehmann et al., 2015, 2016), general surgery (Guerlain et al., 2004; Balayla et al., 2012), endovascular surgery (Brenner et al., 2014), general medicine (Cheung et al., 2018), and dentistry (Aragon & Zibrowski, 2008). The educational strategies that the included studies used were video-based strategies (Chapman et al., 1994; Guerlain et al., 2004; Aragon & Zibrowski, 2008; Lehmann et al., 2016; Cheung et al., 2018), simulation (Aragon & Zibrowski, 2008; Lammers et al., 2008; Brenner et al., 2014; Lehmann et al., 2015; Cheung et al., 2018), lectures (Chapman et al., 1994; Aragon & Zibrowski, 2008), bedside teaching (Lehmann et al., 2016) and interviews (Balayla et al., 2012). Five studies included only undergraduate students as participants (Guerlain et al., 2004; Aragon & Zibrowski, 2008; Lehmann et al., 2015, 2016; Cheung et al., 2018) and one study only postgraduate students (Lammers et al., 2008); two studies included both undergraduate and postgraduate students (Chapman et al., 1994; Balayla et al., 2012); and one included only faculty members who were novices performing the procedure (Brenner et al., 2014). Two studies focused on collecting validity evidence for their instrument (Chapman et al.,

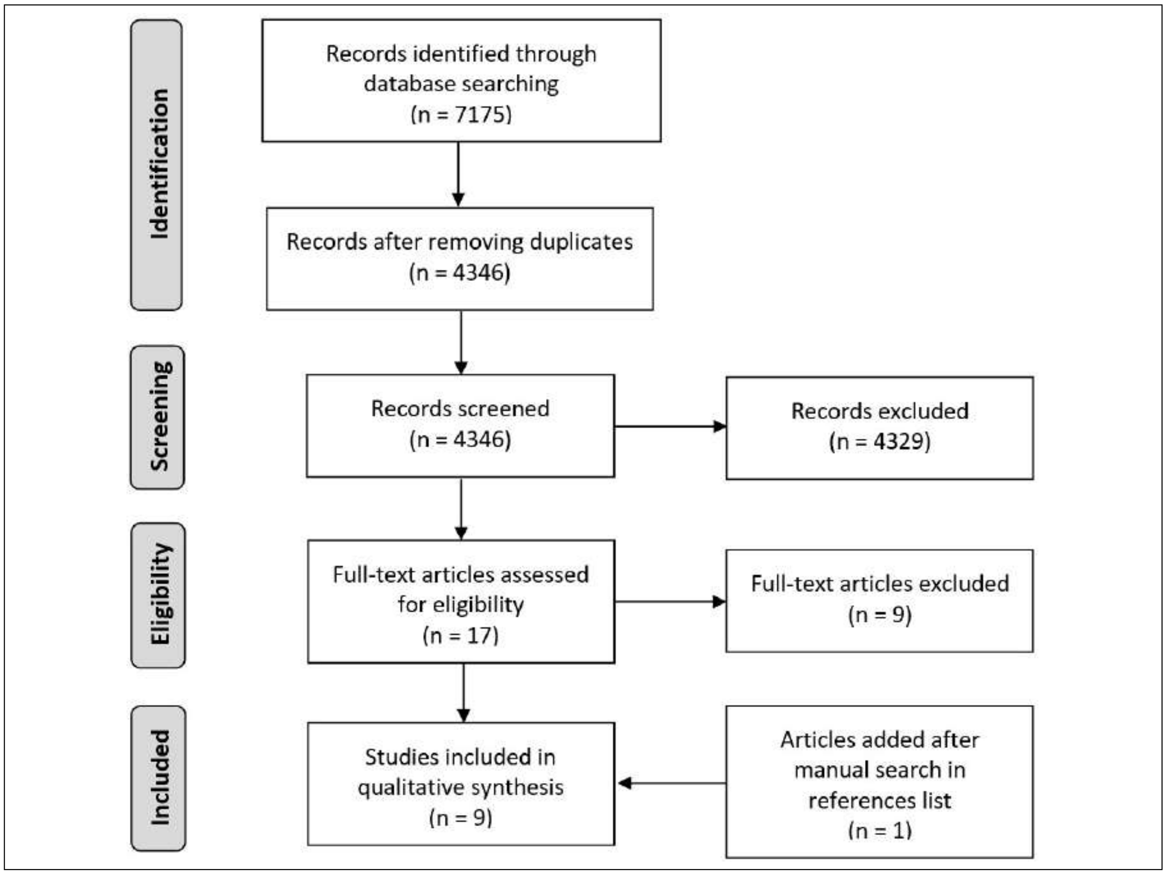


Figure 2.1. PRISMA flow diagram (Moher et al., 2009) illustrating the study selection process.

1994; Balayla et al., 2012), and seven studies on testing an instructional strategy (Guerlain et al., 2004; Aragon & Zibrowski, 2008; Lammers et al., 2008; Lehmann et al., 2015, 2016; Brenner et al., 2014; Cheung et al., 2018). Two of the latter type of studies taught the sequence of steps explicitly and also assessed it (Guerlain et al., 2004; Lammers et al., 2008), two studies taught the sequence of steps explicitly but did not assess it (Aragon & Zibrowski, 2008; Lehmann et al., 2016) and three studies assessed the sequence of steps without explicitly teaching it (Brenner et al., 2014; Lehmann et al., 2015; Cheung et al., 2018). MERSQI median score and interquartile range for the nine studies included was median = 11.5, Q1 = 10, Q3 = 13.5.

Table 2.1. Overview of included articles.

Author	Speciality	Procedure	Instructional Modality	Study Design	Participants	MERSQI
Aragon and Zibrowski (2008)	Dentistry	All-ceramic crown preparation, full gold crown preparation, and posterior porcelain-fused-to-metal fixed partial denture (the 3 procedures with provisional restoration)	Lectures, simulation on mannequin, videos	Case-control design	55 undergraduate students	9.5
Balayla et al. (2012)	General surgery	Inguinal hernia repair with mesh in men. Laparoscopic cholecystectomy and open right hemicolectomy	Interview to validate the assessment instrument	Cross-sectional	35 subjects ('Novice' group: undergraduate medical students, residents in first and second postgraduate year (PGY1-2). 'Expert' group: residents in third to fifth postgraduate year (PGY3-5) and program staff.)	11.5
Brenner et al. (2014)	Endovascular surgery	Resuscitative Endovascular Balloon Occlusion of the Aorta (REBOA)	Virtual Reality Simulation (VRS)	Pre-post test	13 faculty members who are novice interventionalists.	10
Chapman et al. (1994)	Emergency Medicine	Opening the chest, pericardiotomy and aortic cross-clamping	Lectures, videos	Sequential post-test	18 subjects (6 undergraduate medical students, 6 fourth postgraduate year (PGY4), 6 faculty)	14
Cheung et al. (2018)	General Medicine	Lumber Puncture	Simulation in part-task simulator, videos	Pre-post test	29 undergraduate preclerkship medical students (14 control, 15 intervention)	15.5
Guerlain et al. (2004)	General Surgery	Laparoscopic Cholecystectomy	Videos	Pre-post test	30 undergraduate medicine students (14 control, 16 intervention)	12.5
Lammers et al. (2008)	Emergency Medicine	Posterior epistaxis management	Simulation in a model	Prospective, repeated-measures	13 first postgraduate year (PGY-1) residents, 13 second postgraduate year (PGY-2) residents, and 2 third postgraduate year (PGY-3) residents (15 control, 13 intervention)	10.5
Lehmann et al. (2015)	Pediatry	Pediatric Basic Life Support (PBLS)	Simulation on a mannequin	Randomized trial	57 undergraduate medical students (30 control, 27 intervention)	13.5
Lehmann et al. (2016)	Pediatry	Physical examination procedures	Videos, bedside teaching	Survey after intervention	175 undergraduate medical students	9.5

2.3.3. Strategies to teach the standard sequence of steps

Table 2.2 shows the strategies found in each study. Three studies reported the use of videos to demonstrate the standard sequence of steps as a strategy (Guerlain et al., 2004; Aragon &

Zibrowski, 2008; Lehmann et al., 2016). In Lehmann et al. (2016), they used videos with the steps logically ordered in parallel with bedside teaching, while in Guerlain et al. (2004) they used only videos repeating each step three times before moving on to the next step. Aragon and Zibrowski (2008) showed a video with the step-by-step of the procedure during the class, and they gave the students a DVD with the video so they could review it whenever they wanted. Another strategy was used by Lammers et al. (2008) which consisted of informing students of the steps executed in the wrong sequence as soon as they made these mistakes while performing the procedure on a model.

Table 2.2. Teaching the standard sequence of steps: strategies, outcomes and effectiveness.

Authors	**Instructional strategy**	**Outcome**	**Effectiveness**
Aragon and Zibrowski (2008)	Video demonstration	Grades obtained in practical exam (evaluators rate students using a twenty-eight items instrument).	Effective only for one of the three procedures analyzed.*
Guerlain et al. (2004)	Video demonstration	Adherence to the standard sequence of steps (participants answered questions about the sequence of steps).	Effective, it improved performance in procedural questions.*
Lammers et al. (2008)	Informing subjects of all performance and sequence errors immediately	Adherence to the standard sequence of steps (evaluators counted the number of sequence errors).	There were no significant differences between the control and experimental group.
Lehmann et al. (2016)	Video demonstration	Benefits for learning (participants answered an open question).	Videos showed a concrete and standard sequence of steps.

* Statements of Aragon and Zibrowski (2008) and Guerlain et al. (2004) were tested using the corresponding statistic test.

2.3.4. Strategies to assess the sequence of steps' learning

Table 2.3 shows the strategies found in each study. In four studies the authors reported that the sequence of steps was assessed by asking participants to perform the procedure (Chapman et al., 1994; Lammers et al., 2008; Brenner et al., 2014; Lehmann et al., 2015). In Brenner et al. (2014) the evaluators subjectively rated the participants' performance in a virtual reality simulator. In Chapman et al. (1994), the participants performed the procedure in a computer simulation scenario. Evaluators in the study conducted by Lammers et al. (2008) asked the participants to perform the procedure on a model. Similarly, in Lehmann et al. (2015), evaluators assessed the participants by asking them to perform the procedure on a mannequin.

Table 2.3. Assessing the sequence of steps: strategy, outcomes and validity of instruments.

Authors	Strategy	Outcome	Effectiveness
Balayla et al. (2012)	Participants say the steps aloud.	Omissions and commissions (evaluators counted these errors and discounted them of checklist's total score).	Content, response and consequential validity.
Brenner et al. (2014)	Performing the procedure in a virtual simulator.	Adherence to the standard sequence of steps (evaluators used a 5-point Likert scale).	None.
Chapman et al. (1994)	· Students write the steps. · Performing the procedure in computer simulation. · Performing the procedure on an animal model, and saying the steps aloud.	Adherence to the standard sequence of steps, omissions and commissions (both outcomes were evaluated assigning a score to each step using a rating scale).	Content validity for all assessment strategies. Internal structure validity for animal and computer assessment.
Cheung et al. (2018)	Participants write the steps in the proper sequence.	Adherence to the standard sequence of steps (evaluators assigned points to each participant).	None.
Guerlain et al. (2004)	Students answer a test with questions about the sequence of steps.	Adherence to the standard sequence of steps.	None.
Lammers et al. (2008)	Performing the procedure on a model.	Adherence to the standard sequence of steps (evaluators counted the number of sequence errors).	Response validity.
Lehmann et al. (2015)	Performing the procedure on a mannequin.	Adherence to the standard sequence of steps (evaluators rated each step through a rating scale, considering whether the step was omitted, done in the correct or incorrect position).	Content and response validity.

In two studies participants were asked to say their actions aloud (Chapman et al., 1994; Balayla et al., 2012). In Balayla et al. (2012), they were asked to say the steps in the sequence they remembered them, while in Chapman et al. (1994), they had to verbalize the steps while performing the procedure on an animal model. Two other studies asked participants to write the procedure's steps in the proper sequence (Chapman et al., 1994; Cheung et al., 2018), and in another study the participants answered three true/false and multiple-choice questions about the sequence of steps (Guerlain et al., 2004).

2.3.5. Instruments to assess the sequence of steps' learning

The evaluators used different instruments with different rating scales to assess the sequence of steps. In Brenner et al. (2014) they used a 5-point Likert scale. In Cheung et al. (2018), they assigned 13 points to the sequence of steps written by the students (the procedure has 13 steps). In Lehmann et al. (2015), they assigned a score to each step considering whether it was done in the correct position of the sequence: 2 points if it was done in the correct position; 1 point if it was done in the wrong position, or 0 points if it was not done. In Chapman et al. (1994), they considered adherence to the standard sequence of steps as one of four items to assign a score to each step. On the other hand, some studies did not use rating scales. In Guerlain et al. (2004), evaluators used true/false and multiple-choice questions about the sequence of steps. In Lammers et al. (2008), they counted the number of sequence errors that the participants committed.

2.3.6. Outcomes to assess the sequence of steps' learning

Table 2.3 shows the outcomes found in each study. Six studies measured the adherence to a standard sequence of steps (Chapman et al., 1994; Guerlain et al., 2004; Lammers et al., 2008; Brenner et al., 2014; Lehmann et al., 2015; Cheung et al., 2018). To measure this outcome, the evaluators used the instruments mentioned in the previous section. Also, two studies measured the number of omissions and commissions committed (Chapman et al., 1994; Balayla et al., 2012). The omission of steps refers to missing them, and commission refers to unnecessarily adding or doing steps in the wrong sequence (Balayla et al., 2012). In Balayla et al. (2012), they used this outcome to penalise the checklist's total score with the number of omissions and commissions made. On the other hand, in Chapman et al. (1994), they considered both errors as one of four items to assign scores to each step.

2.3.7. Validity of instruments to assess the sequence of steps

Table 2.3 shows the evidence found in each article. Three studies used assessment instruments without validity evidence to assess the sequence of steps (Guerlain et al., 2004; Brenner et al., 2014; Cheung et al., 2018) , i.e., experts in the procedure involved were not asked for their opinion on the instrument, nor was the consistency of these instruments evaluated when assessing residents. Three studies presented evidence for content validity (Chapman et al., 1994; Balayla et al., 2012; Lehmann et al., 2015) since experts were asked about the instrument's suitability. One study presented evidence for internal structure validity (specifically in the animal model and computer assessment) (Chapman et al., 1994), which means that such instruments contain different questions to assess the same skill. Three studies presented evidence for response validity (Lammers et al., 2008; Balayla et al., 2012; Lehmann et al., 2016) showing the use of blinded raters, and only one study presented evidence for consequential validity (Balayla et al., 2012) deciding the approval or fail of the training course with the instrument under analysis.

2.3.8. Effectiveness of strategies to teach the standard sequence of steps

Table 2.2 shows the outcomes and effectiveness of the strategies found. In Guerlain et al. (2004), the results showed that the students' performance in the true/false and multiple-choice questions improved after the intervention. Lammers et al. (2008) found no significant differences between control and experimental groups when pointing out to students the sequence errors they committed. In Lehmann et al. (2016), they asked the students about the benefits of videos for learning and self-confidence, and the most frequent answer was that videos showed them a concrete and standardized sequence of steps. Finally, Aragon and Zibrowski (2008) measured the grades of the participants in a final course test and found a positive correlation between the use of the videos and the grades, but they only found it in one of the three procedures (all-ceramic crown preparation and provisional restoration procedure). When comparing the grades, they found that the intervention participants obtained higher test scores than those of the control group, but only for the procedure mentioned.

2.4. Discussion

In this systematic review, we searched the literature for studies reporting strategies to teach and assess the sequence of steps' learning in procedural skills training. The results show that the teaching of a standard sequence of steps and the assessment of this aspect is rarely reported in procedural skills training studies. Also, studies' quality does not allow to determine whether teaching a standard sequence of steps or assessing it explicitly have a positive impact on learning this aspect. The latter refers to the lack of validity evidence for instruments to explicitly assess the standard sequence of steps' learning (Chapman et al., 1994; Lammers et al., 2008; Lehmann et al., 2015), and the need to optimize teaching strategies designed for this aim (Guerlain et al., 2004; Aragon & Zibrowski, 2008; Lammers et al., 2008; Cheung et al., 2018).

Only nine studies had reported the standard sequence of steps' inclusion in their didactic methods. Most of the studies already in the literature focus on teaching and assessing each step separately, and also in decomposing a procedure in steps through cognitive task analysis (Clark et al., 2012). They reported a variety of strategies to teach and assess the sequence of steps explicitly. Regarding the quality of studies, the MERSQI median score of the nine studies included is moderate (Reed et al., 2007). Also, studies designed as a randomized controlled trial tended to have a greater overall MERSQI score. Furthermore, most of the included studies had focused on the second Kirkpatrick level (knowledge, skills) (Kirkpatrick, 2007).

Three studies included in this review reported the use of videos to teach the standard sequence of steps (Guerlain et al., 2004; Aragon & Zibrowski, 2008; Lehmann et al., 2016). These studies reported its effectiveness for learning, except in the study conducted by Aragon and Zibrowski (2008), where learning improved only in one of the three procedures they taught. In a recent review, Green et al. (2019) found that in most of the interventions that employed videos, the learning of the participants improved. Also, they mentioned that videos are useful to expose students to procedures that are rarely performed (as is the case of dentistry) (Aragon & Zibrowski, 2008) and they recommend the use of additional material to the videos for their effective use, such as a narration of the procedure or a diagram that represents it.

Another strategy used to teach the standard sequence of steps was simulation-based feedback. The use of simulation allows training in a safe environment without harming patients (Borgersen et al., 2018), it is effective as a learning modality of procedural skills (Cook et al., 2013), and in some studies had proven to be cost-effective (Zendejas et al., 2013). However, the immediate feedback used by Lammers et al. (2008) did not show significant differences between both groups. Despite this result, this strategy prevents students from keeping the errors in their long-term memory, and thus students are less likely to commit the same errors in the future (Nicholls et al., 2016).

To determine the effectiveness of an instructional strategy, it is recommended to have an adequate alignment between the teaching strategy and the assessment task to evaluate skill's learning (L. W. Anderson et al., 2001). Two included studies presented some level of alignment between the teaching strategy reported and the assessment task, one of them had positive effects (Guerlain et al., 2004) while the other did not (Lammers et al., 2008). Hence, the effectiveness of strategies to teach the standard sequence of steps explicitly remains unclear.

Six studies assessed adherence to a standard sequence of steps (Chapman et al., 1994; Guerlain et al., 2004; Lammers et al., 2008; Brenner et al., 2014; Lehmann et al., 2015; Cheung et al., 2018). They used a variety of instruments to measure this outcome (Likert scale, different scoring protocols, multiple-choice, and true/false questions). On the other hand, two studies explicitly measured and incorporated the omission and commission of steps into their assessment instruments (Chapman et al., 1994; Balayla et al., 2012). Some authors perceived that checklists do not usually allow assessing adherence to a standard sequence of steps, omission, or commission of steps (Lammers et al., 2008; Lehmann et al., 2015). This review supports this perception because we found few studies measuring these outcomes. Therefore, further research might help to develop assessment instruments that consider the omissions, commissions and adherence to a standard sequence of steps, thus assessing the sequence of steps during procedural skills training with instruments suitable for this purpose.

Regarding the validity of the instruments collected in this review, four studies presented some type of validity (Chapman et al., 1994; Lammers et al., 2008; Balayla et al., 2012; Lehmann et al., 2015), while three studies did not present evidence of validity (Guerlain et al., 2004; Brenner et al., 2014; Cheung et al., 2018). The lack of instruments' validity to explicitly assess the sequence of steps is a research gap to be addressed (Chapman et al., 1994; Lammers et al., 2008; Lehmann et al., 2015), which can be covered using contemporary validity frameworks, following the recommendation of Borgersen et al. (2018).

Teaching a standard sequence of steps and assessing its learning is crucial to prepare residents to deal with procedural variability (Apramian et al., 2015, 2016; Légaré et al., 2015). This incorporation will improve the residents' training and prepare them better for future challenges (Apramian et al., 2016). However, it has been seen that instructors struggle to guide residents on what is a principle and what is a preference, and they typically are not explicit about the procedural variability (Sullivan et al., 2014; Apramian et al., 2015). We propose the use of Surgical Process Models to explicit the possible variations in the sequence of steps of a procedure, and also to differentiate between principles and preferences.

Surgical Process Models (SPM) is a recent area dedicated to modeling surgical procedures (Lalys and Jannin, 2014; Neumuth, 2017). SPM models are 'a simplified (formal or semiformal) representation of a network of surgical or surgery-related strategies and their relationships' (Neumuth,

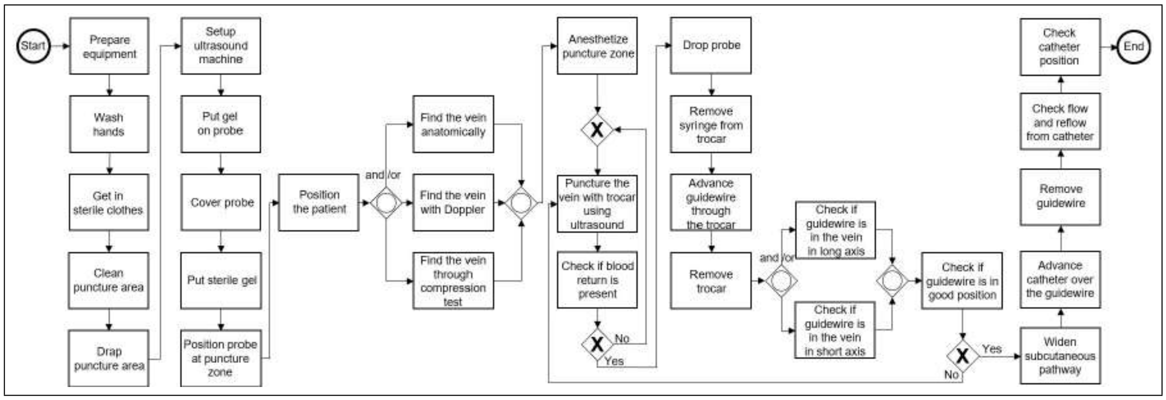


Figure 2.2. Surgical Process Model depicting the central venous catheter installation procedure (based on de la Fuente, Fuentes, Munoz-Gama, Dagnino, and Sepúlveda (2020)).

2017), and this simplification allows to understand surgical procedures as a collection of steps that are sequentially ordered (Lalys and Jannin, 2014; Neumuth, 2017). Furthermore, the formality of the modeling languages on which SPM models are built allows to visualize decision points along the procedure, clearly defining what is a principle and what is a preference. Figure 2.2 shows a Surgical Process Model of the central venous catheter installation procedure (de la Fuente et al., 2020). This model makes explicit the sequence of steps to perform this procedure, the decision points where it is possible to choose between steps representing different options, and where parts of the procedure could be repeated.

Surgical Process Models might help to report the procedure's standard sequence of steps in research studies, to compare surgical approaches, to explicitly teach the standard sequence of steps and assess the adherence to it in procedural skills training (Lalys & Jannin, 2014; Neumuth, 2017; Vedula & Hager, 2017). Furthermore, these models would help to make the sequence of steps explicit as a learning objective, producing the desired alignment between objective, teaching strategy and assessment (L. W. Anderson et al., 2001).

Instructors can use an SPM as a procedural diagram to depict the procedure's standard sequence of steps. An SPM resembles a flowchart of the procedure that shows the steps and their standard sequence of execution (de la Fuente et al., 2020). Teaching a surgical procedure using an SPM might help students visualize the sequence of steps, focus the training on the sequence of steps, visualizing what is a principle and what is a preference, rehearse the sequence of steps in the parts of the procedure that were difficult to perform (Rao et al., 2015; Wallace et al., 2017) and provide feedback focusing on the sequence of steps. These strategies could complement the information provided by videos or the instructor of the training. The strategies mentioned in the prior sentences are relevant because instructors omit about 70% of the information that students need during their learning process (Sullivan et al., 2014), and it is difficult for experienced surgeons to share their mental models (Crebbin et al., 2013; Apramian et al., 2015; Lateef, 2018). Additionally, these strategies would serve to simulate real situations that rarely occur and expose students unusually performed procedures (Osterweil et al., 2019).

It is important to mention that, despite the potential benefits that Surgical Process Models bring to procedural skills training, implementing and creating them is not straightforward. One reason is that SPMs need to be comprehensive, i.e., to represent the most of the procedural variations that a surgical procedure has, thus making the model suitable to different patients, surgeons and hospital resources (Rojas et al., 2016; Neumuth, 2017). However, to address this issue, the variations that residents need to know at the end of the training (e.g., in a competency-based training) can be selected through consensus, and thus including them to create a suitable SPM. A second reason is that it remains unclear whether following the sequence defined by SPMs ensures positive patient outcomes, and further research is needed to analyze the impact of following the sequence defined by a SPM on outcomes.

A limitation of this review is that we did not analyze the effectiveness of the teaching strategies described through meta-analysis since they all used different ways of measuring it and the misalignment found between teaching and assessment strategies. Another limitation is that the effectiveness of the sequence of steps' teaching strategies described could be biased since most of the studies found did not have the sequence of steps as the unique focus.

2.5. Conclusions

The standard sequence of steps is an aspect rarely reported in procedural skills training studies. The included studies presented high variability in the strategies and instruments used to explicitly teach and assess the learning of a standard sequence of steps. Also, the studies' quality prevents determining whether the strategies for teaching or assessing the sequence of steps have a positive impact on the learning of this aspect. Therefore, more research is needed to find methods and strategies that ensure the learning of this aspect during procedural training and, consequently, prepare residents to deal with procedural variability better. Using innovations such as Surgical Process Models might enable the design of new strategies and instruments to incorporate the standard sequence of steps in procedural skills training.

Chapter 3

Interactive Process Mining for Medical Training

3. INTERACTIVE PROCESS MINING FOR MEDICAL TRAINING

Process Mining has been widely used in healthcare in different medical areas (Rojas et al., 2016), and recently some applications in the medical training field have been developed. In particular, the use of Process Mining in the training of procedural skills has opened a branch of opportunities to fill gaps in this field.

Procedural skills are essential to perform surgical procedures and to obtain good clinical outcomes (Fecso et al., 2017). Literature suggests that surgical procedures can be seen as a process (Neumuth et al., 2011), so it is possible to analyze surgical procedures with Process Mining. This perspective allows focusing on the sequence of steps of a surgical procedure, an aspect rarely considered in the medical training research and practice. Also, this view enables the development of different applications that can be useful in the medical training context for tasks like teaching, assessment, giving feedback, among others.

In medical education, videos are regularly used as a tool in surgical teaching (Green et al., 2019). However, it is not clear how to use them and the effectiveness of the approaches in which videos are involved. Additionally, it is required to not interfere with the behavior of novices performing the procedure, to collect reliable data and avoid bias in it. It is also needed a method that considers time flexibility, since doctors are under pressures and constraints to dedicate time to their research and teaching duties (Roshetsky et al., 2013). Therefore, a method to collect process data from videos containing a procedural skill performance is essential to enable the research and practice of the process-oriented perspective of procedural skills.

3.1. Running case: Central Venous Catheter insertion

In this chapter, we will use the Central Venous Catheter insertion as a running case to illustrate the POME (Process-Oriented Medical Education) method, which is a common surgical procedure performed by anesthesiologists and intensivists. This procedure has six main steps: first, prepare implements and patient for the procedure; then, a vein is punctured using a trocar (a needle with a hole to introduce a guidewire); next, the guidewire is passed through the trocar; later, the trocar is removed, and the catheter advanced through the guidewire; finally, the guidewire is removed and the catheter installed.

3.2. POME method

Figure 3.1 shows the POME (Process-Oriented Medical Education) method overview. This method facilitates the analysis of surgical procedures as a sequence of steps and uses the results for medical training tasks. It is composed of three stages: first “Model Stage”, second “Data Stage” and third “Analysis Stage”. Each stage has its components and relations between them, which we explain below.

The “Model Stage” consists of developing a graphical representation (i.e. a model) of the surgical procedure as a process. In the “Process Modeling” step, a first draft of the model is designed and is then assessed the model experts agreement level through a “Delphi Panel” step. Experts should be doctors who have experience performing the procedure. Both steps are iterative: depending on the level of agreement reached in the “Delphi Panel” the model is modified, to then assess the expert’s agreement level with the model again. This stage ends when the level of agreement reached is the desired.

The “Data Stage” focus on generating data to analyze surgical procedures as processes. That means creating Event Logs. In order to do so, executions of the procedure are needed, which are commonly captured through video recordings (“Execution and Recording” step). These videos are used for different tasks in medical education, but still it is not clear their effectiveness and how to use them (Green et al., 2019). We tag the videos with the activities defined in the model developed in “Model Stage”. Tagging videos allow getting the entire sequence of steps of an execution, and therefore an Event Log with all the executions.

In the “Analysis Stage” the Event Log generated in “Data Stage” and the model generated in “Model Stage” are used to perform the analysis with Process Mining algorithms. With the information obtained after treating the data, it is possible to create a report (“Process Reporting” step). Designing the report will depend on the goal of the application. However, the main requirement is to create an easy-to-interpret report for doctors.

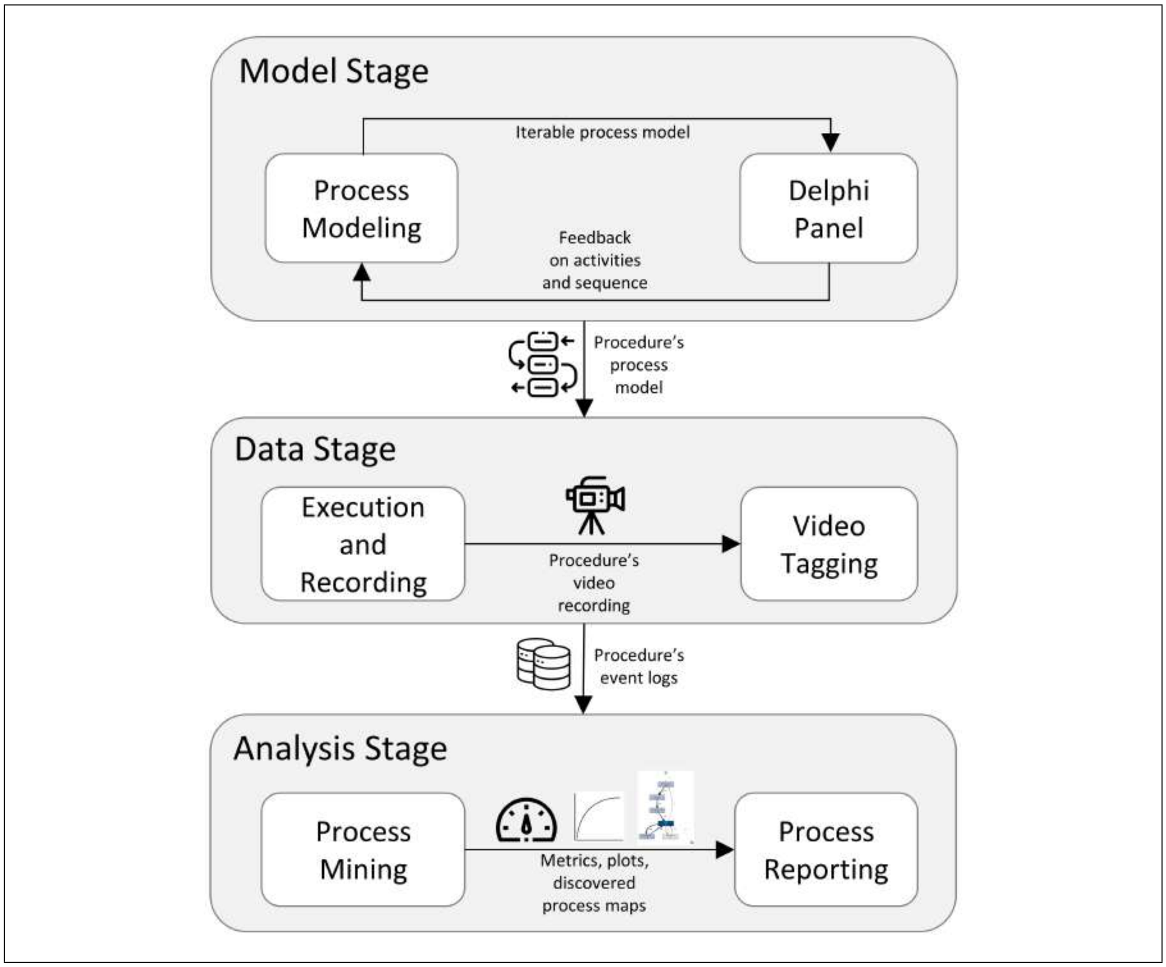


Figure 3.1. POME method overview.

3.3. Model Stage

3.3.1. Process Modeling

The first step in our method is to have a Generic Surgical Process Model of the procedure under analysis. This model has been defined as “a simplified pattern of a medical procedure in a formal or semi-formal representation” (Neumuth et al., 2011). This generic model will not only be useful to have a reference standard to compare the executions made; rather, it is a representation in which the procedure is broken down into sequential steps, decision points and alternative pathways.

This breakdown of the procedure has been defined as an input that all procedural training must have (Grantcharov & Reznick, 2008; Wingfield et al., 2015; Huang et al., 2016).

The development of this model is not trivial due to the inherent difficulties of generating process models for the healthcare and medical education domains:

(i) Surgical processes show a lot of variability among executions due to the experiences, skills and preferences of the health personnel, the patient's characteristics, and the availability of resources and technology (Müller and Rogge-Solti, 2011; Neumuth et al., 2011).

(ii) When consulted how they perform a procedure, experts tend to omit relevant information (Yates et al., 2012). This omission can reach up to 70% of the steps necessary for a correct execution (Sullivan et al., 2014), and it has been attributed to the automation of high levels of expertise (Hoffman, 2008).

(iii) An adequate representation of a procedure requires a holistic approach to procedural competence. Whoever performs a procedure must have not only the necessary technical skills, but also the skills that ensure the patient's physical and psychological comfort, such as the care necessary to avoid mechanical and infectious complications. Thus, McKinley et al. (2008) have defined the following dimensions as necessary components of any representation of a procedure: preparation; infection control; communication and work with the patient; teamwork; security; procedural competence; post-procedure care (McKinley et al., 2008).

In this context, the objective when developing a model for Process-Oriented Medical Education (POME) is to have a model without local or specialty biases, versatile to be applied to different settings and centers, complete from the point of view of having all the technical information necessary for the execution and complete because it includes those steps necessary to obtain a holistic representation of the competencies required for adequate health care.

Thus, the generic process model is obtained in two stages: one is the generation of a first model of the procedure, and second is the validation of the model using the Delphi Methodology. For the first stage, we rely on published checklists for the chosen procedure. A checklist is a list of observable activities or behaviors, organized consistently, that allows an observer to record the performance dichotomously (i.e. done or not) in an assessment context (Hales et al., 2008). All published checklists are analyzed in terms of their psychometric validity with support of validation frameworks, their completeness of activities ensuring the presence of all the activities contained in the checklists collected, and the presence of the seven dimensions of competence defined by McKinley et al. (2008) through domain knowledge. Using the list of activities defined in the checklists as a reference, a representation of the procedure is constructed in BPMN notation, a notation that, in addition to being a de facto standard for process modeling, has proven to be easily understood by users in the healthcare area (Scheuerlein et al., 2012; Rolón et al., 2015). The result of this first stage is a first generic process model, which will be subjected to a validation process that avoids biases of specialty or local practices that make the model little applicable to other health centers or realities. The process is explained below.

3.3.2. Delphi Panel

Delphi methodology has proven to be an effective tool in many disciplines to achieve consensus among experts on a given topic (Hasson et al., 2000; Diamond et al., 2014). It is characterized by the anonymous interaction of experts, who in successive and controlled rounds can modify their answers after knowing the answers of the rest of the participants. This interaction concludes when the consolidated responses represent the majority of the group (Mead & Moseley, 2001). The realization of the Delphi panel requires a structured characterization and selection of experts. For this purpose, we used the recommendations of Okoli and Pawlowski (2004). In our case we define a minimum time of experience in the procedure, a minimum number of monthly executions and additionally meet one of the following characteristics: be the local manager of the procedure, be an accredited instructor, be the head of a service where the procedure is performed frequently or have participated in guidelines or publications regarding the performance of the procedure.

Once the experts from different specialties and health institutions have been selected, they are invited to participate in an online survey. In the survey, the activities defined in the first model are ordered sequentially, asking the experts to express their agreement with the inclusion of this activity in the final model, through a 5-point Likert scale: (1) under no circumstances should be included, (2) should not be included, (3) may or may not be included, (4) should be included, (5) must be included. Also, they are asked to propose new activities, modify proposed activities, and propose changes to the place they should occupy in the sequence. Once the experts complete the survey, the results obtained for each activity are presented in a second survey, showing them as the percentage obtained by each item on the Likert scale. Also, the new proposed activities are added, and they are asked to express themselves regarding the suggested modifications for any activity. In this second survey, they are again invited to weigh the inclusion of each activity in the final model of the procedure based on the same 5-point Likert scale. This sequence is repeated up to a third time if the previously defined agreement criterion is not reached. To ensure the adequate methodological quality, planning and execution of the Delphi panel, it should follow the recommendations of Diamond et al. (2014): a reproducible selection of participants, the definition of a stopping criterion, a maximum number of rounds, and an exclusion criterion for each item (Diamond et al., 2014).

The proposed modeling methodology allows obtaining a representation of the procedure in BPMN notation, based on the information available in publications and subsequently enriched through the consensus of experts from different centers. This mixture of information allows us to have a process model without local or specialty biases that can be applied to analyze.

3.3.3. Running case model

After conducting the model stage, we obtained a BPMN process model representing the central venous catheter installation. Figure 3.2 shows the model obtained for the running case, and the details on how it was generated (such as panel size, level of agreements, other relevant information) can be found in de la Fuente et al. (2020).

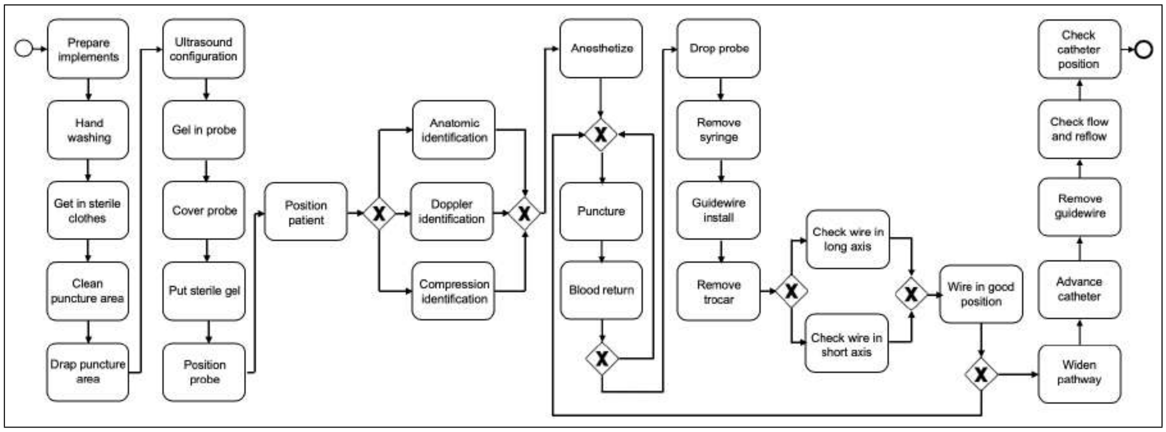


Figure 3.2. BPMN model of the Central Venous Catheter installation (adapted from Munoz-Gama et al. (2019)).

3.4. Data Stage

3.4.1. Execution and Recording

Processes analyzed with Process Mining commonly have an information system behind them, recording all the data generated during their execution. Even when its database is not recording the data with an Event Log shape, it is possible to build them using this raw data. In Leiva et al. (2019), these type of processes are called *plugged* processes (left image in Fig. 3.3).

However, some processes are not supported by information systems, because some parts of its execution are not recorded in common databases, are based on manual work or involve the mixture of other data sources than common database systems (e.g. paper data or logs, spoken decisions). Process Mining can help for analyzing these processes, but creating the Event Log needs a different treatment than *plugged* processes. In Leiva et al. (2019), these type of processes are called *unplugged* processes (right image in Fig. 3.3).

How to collect data to analyze *unplugged* processes is the question, and the answer will depend on the context. In the medical training field, the execution of surgical procedures commonly are filmed, so the primary source of data comes from them.

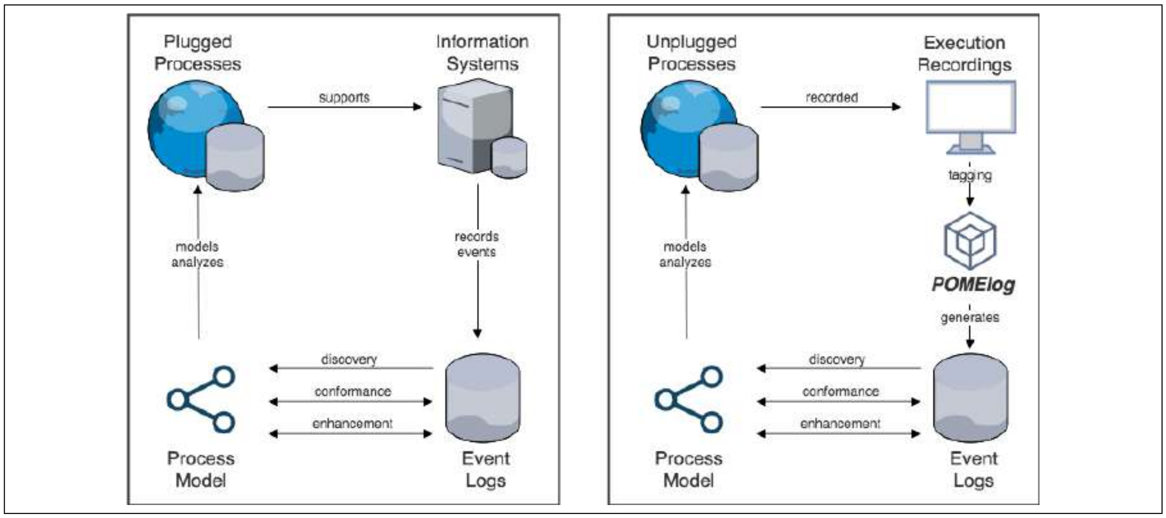


Figure 3.3. Process Mining for plugged processes (left) and unplugged processes (right).

3.4.2. Video Tagging

As we previously mentioned, videos are the main source of data in medical education. However, we need data in the event log shape to perform process mining analyses. Therefore, the POME method involves video tagging as a way to obtain data from videos, which should be done by experts in the surgical procedure involved. Because of that, the lack of surgeons experienced and the little available time they have (Walter, 2006) is a challenge that needs to be addressed. In this step, experienced doctors are needed or, at least, doctors well trained in how to execute the procedure. This is crucial to ensure data quality and unbiased results.

Methods to control the bias generation during data collection, as well as methods to generate the Event Logs will depend on the data resource type used. To avoid bias, in our running case, we use the Levenshtein distance (Cohen et al., 2003) to decide how different are the tags between different taggers. Once the taggers labelled all the videos, we compare the tagging generated by them for each video using the Levenshtein distance. If the taggings are similar, we can use any of them in the analysis stage. Otherwise, we need to decide which of the taggings is the most reliable, typically using a third party to decide on which them to use. To generate the Event Logs, we developed POMElog, a web-based platform where videos can be tagged and doing so generate the data.

POMElog (Leiva et al., 2019) allows doctors to tag videos in a user-friendly way. Following Figure 3.4, POMElog contains all the activities of the model designed in the Model Stage (section A), different views to help the tagger precisely decide which activity is being executed (section B), give the option to select the starting and ending point of time an activity is executed (section C), adjust the speed of the video (section D) and finally export the event log (section E).

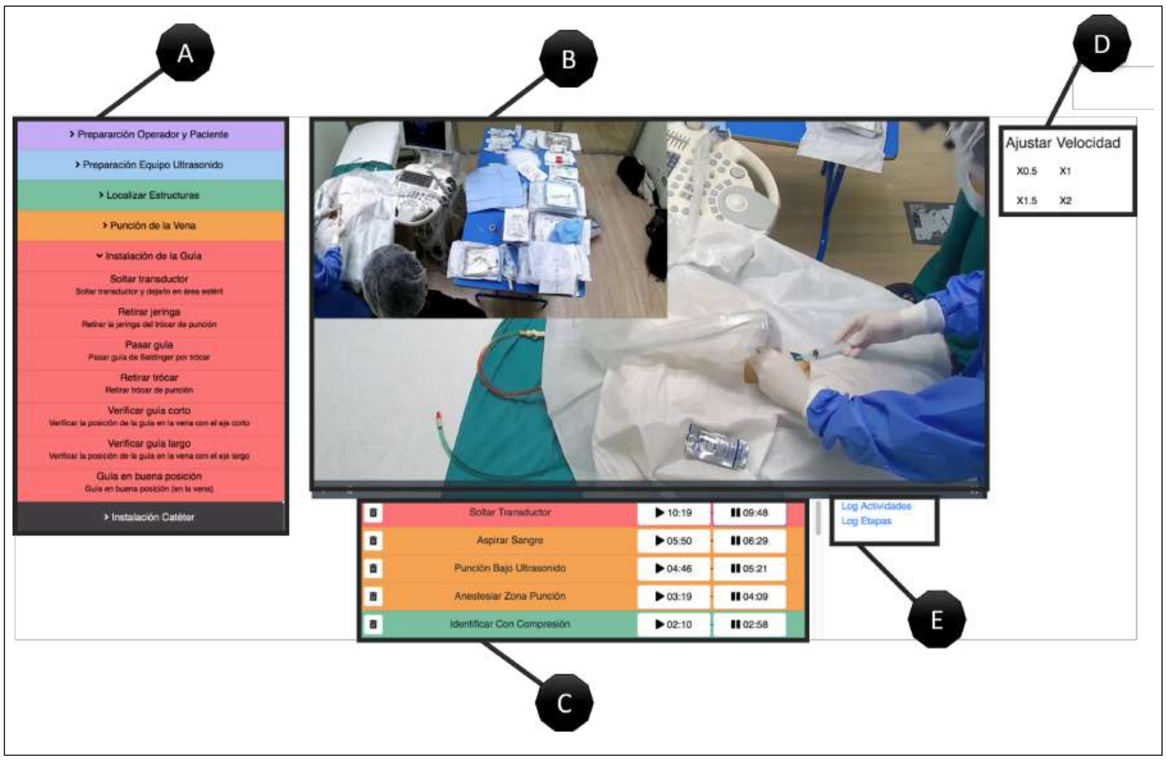


Figure 3.4. POMElog and features. (Leiva et al., 2019)

3.4.3. Running case data

In our running case, we uploaded the videos of residents performing the central venous catheter installation in a simulation environment to POMElog, and this platform delivers the Event Log ready to be analyzed using Process Mining.

3.5. Analysis Stage

3.5.1. Process Mining

Once created the Event Logs, it is time for the "Process Mining" step. Process Mining algorithms receive as input the data, and the chosen algorithm depends on the task of interest. If the objective is to know the common pathway followed by executions of a surgical procedure, Discovery algorithms can be used to see it and its deviations. If the objective is to compare the model generated in the "Model Stage" of POME method with data obtained in "Data Stage", Conformance Checking algorithms can help to accomplish this task.

3.5.2. Process Reporting

After the "Process Mining" step ends, it is necessary to design an easy-to-understand report for doctors. The "Process reporting" step consists of showing the results of the last step in a way doctors can understand and use in medical training tasks. The report will vary depending on the objective for what it was generated.

Designing the reports should consider that doctors are not experts in Process Mining but in healthcare. It is essential to establish requirements for the report, and then test with the medical educators if the report accomplishes them. Also, it is crucial to evaluate their understanding and ease of interpretation of the report. With this, the application use's likelihood will increase. Validation techniques (Wieringa, 2014) can help on this task through techniques such as expert opinion and effects analysis.

3.5.3. Running case analysis

The objective of the analysis of the running case was to give feeback about the sequence of steps to residents learning the central venous catheter installation. After performing a discovery study, a feedback report was generated (Lira et al., 2019) and tested with the residents' opinion to assess its acceptance. Figure 3.5.

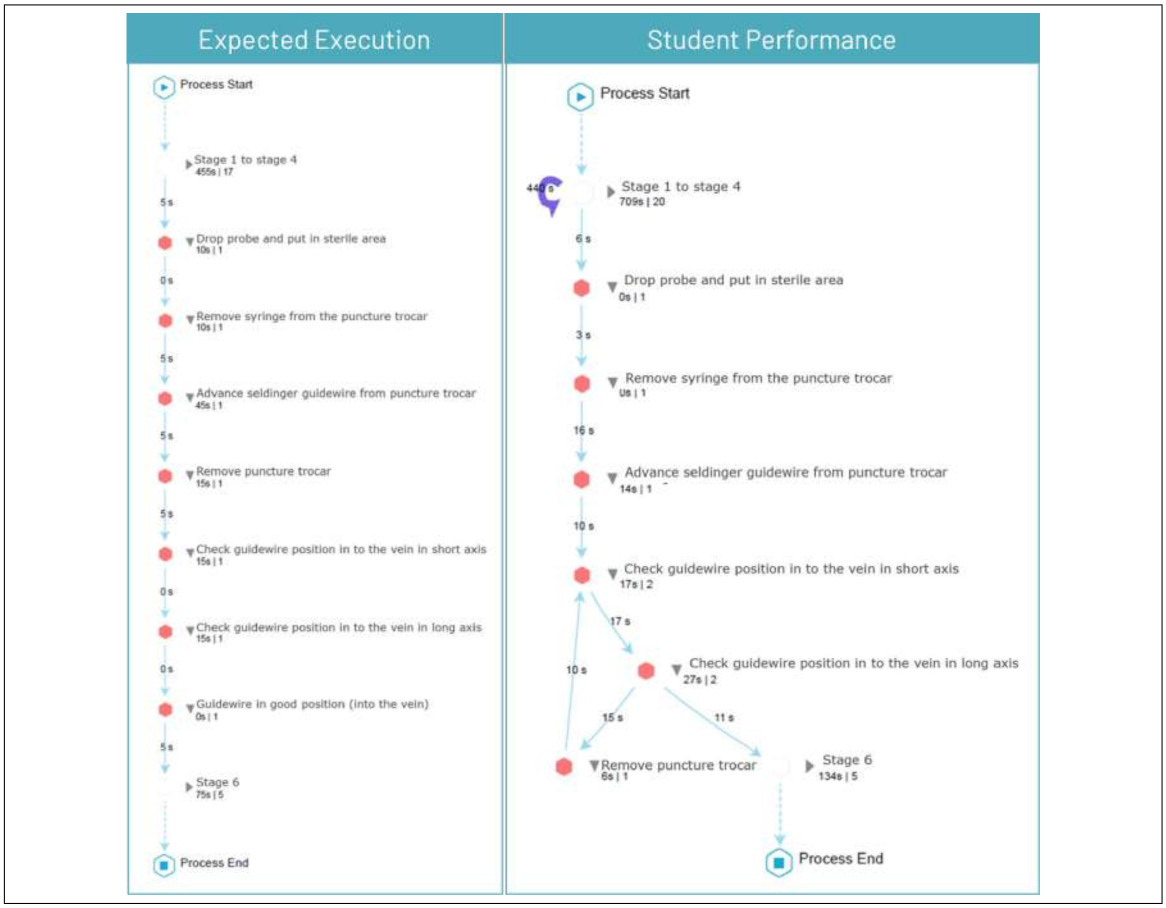


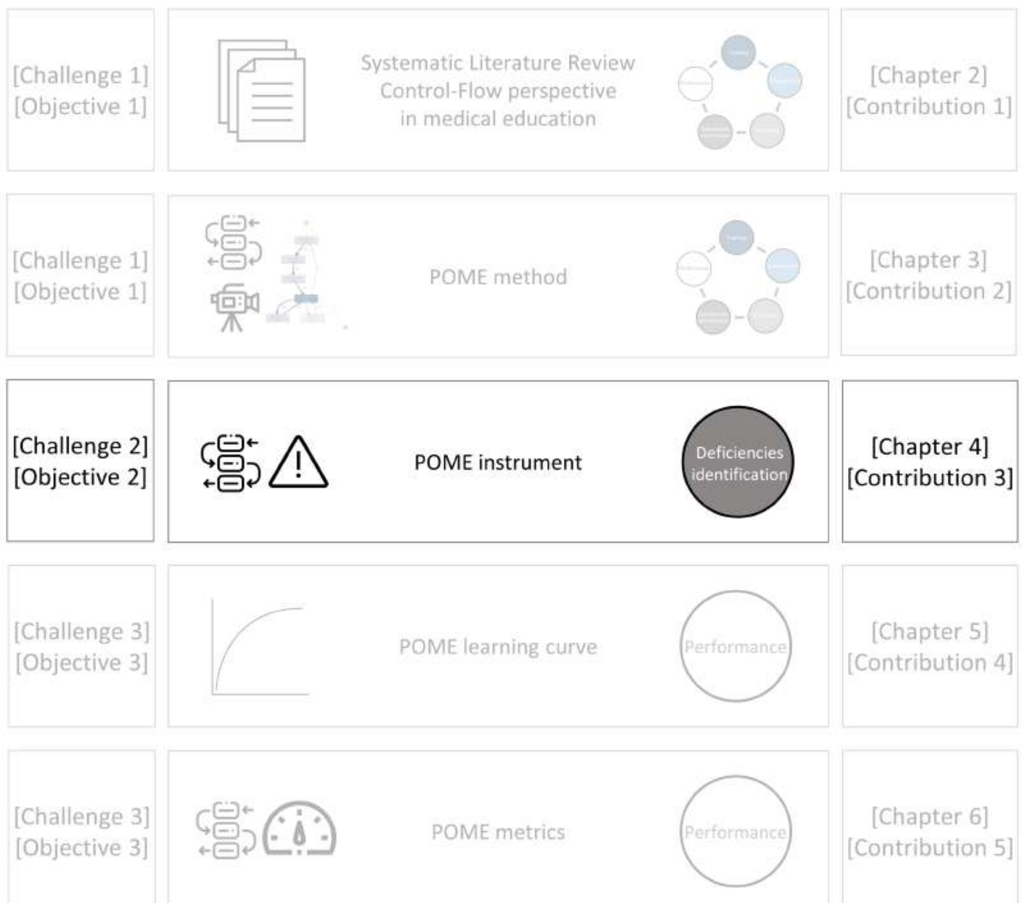
Figure 3.5. Diagram of the Guidewire Install surgical procedure stage included in the feedback report. The expected execution is shown on the left side and the student's performance is shown on the right side (Lira et al., 2019).

3.6. Conclusion

This chapter describes the POME method, which allows considering the sequence of steps as a relevant aspect in medical training applications. This method proposes a novel strategy to analyze surgical procedures as processes, creating all the elements needed to run a Process Mining project. The"Model stage" allows to obtain an abstraction of a surgical procedure, the "Data stage" proposes a new way to use videos and obtain data from them, and the "Analysis Stage" take in consideration context variables as bias and ease-of-understanding of the information given to doctors. We used the Central Venous Catheter insertion procedure as a running case, showing a successful case of using this method. This example encourages the use of Process Mining with other surgical procedures, enabling the development of new tools to the medical training field.

Chapter 4

Process-Oriented Instrument and Taxonomy for Teaching Surgical Procedures in Medical Training: The Ultrasound-Guided Insertion of Central Venous Catheter



4. PROCESS-ORIENTED INSTRUMENT AND TAXONOMY FOR TEACHING SURGICAL PROCEDURES IN MEDICAL TRAINING: THE ULTRASOUND-GUIDED INSERTION OF CENTRAL VENOUS CATHETER

4.1. Introduction

Medical education aims to prepare physicians with the latest scientific discoveries in the prevention and treatment of illnesses and diseases that people suffer (*WMA Statement on Medical Education*, 2017). Surgical procedures are an essential part of these treatments, and doctors need to be proficient in procedural skills to perform them successfully. Procedural skills (i.e., skills needed to perform surgical procedures) are one of the technical competencies considered when teaching (Aydin et al., 2017), and they are relevant because they are associated with good clinical outcomes (Fecso et al., 2017).

An instructor who teaches procedural skills needs to understand student performance regarding the sequence of steps of a surgical procedure. With this in mind, instructors can develop strategies to provide specific feedback and assess their students, which are competencies an instructor should have (Srinivasan et al., 2011).

Also, instructors assess procedural skills using standard tools, like checklists with the steps needed to complete a procedure (Shackelford & Bowyer, 2017), and Global Rating Scales (GRS) (I. W. Y. Ma et al., 2012) to qualitatively assess indicators such as the flow of the procedure and economy of movements (Shackelford & Bowyer, 2017). Tools mentioned above are useful, but state-of-the-art tools rarely take into account the sequence of steps: they are focused on each isolated step and do not consider the relative importance of each step or the incorrect execution of the sequence of steps.

Process Mining (van der Aalst, 2016) is a new discipline that allows the analysis of processes using data stored and generated by information systems that support them. This discipline has been successfully used in a wide variety of healthcare specialties (Rojas et al., 2016), among them medical education. Lira et al. (2019) showed the use of Process Mining to give specific feedback, using data obtained from recorded executions of a procedure performed by students. In addition, de la Fuente et al. (2020) used Process Mining to compare trainees and experts with the ideal sequence of steps. However, neither study emphasized information relevant to instructors, and they are difficult for them to interpret.

Surgical procedures are a progression of steps (Lalys & Jannin, 2014) that can be seen as a process (Neumuth et al., 2011). Therefore, in this research, we used Process Mining to help instructors to understand students' performance regarding the sequence of steps in surgical procedures. The information captured by an analysis of the sequence of steps through Process Mining adds information that is not possible to capture with a checklist, which is limited to a dichotomous assessment of the execution of specific steps of a procedure, without considering the order or the unnecessary repetition. Conversely, it explicitly explains a qualitative dimension included in the GRS, such as fluidity in the execution of the steps of a procedure.

The objectives of this research were: (1) to define a taxonomy of steps for surgical procedures; (2) to design an instrument for instructors with process-oriented information about the sequence of steps followed by their students; and (3) to evaluate with instructors the usefulness of both tools for their tasks as teachers. The approach presented has two steps: first, we developed the taxonomy, and then we designed the process-oriented instrument based on the taxonomy.

This article includes: (1) information about current ways instructors use to teach the sequence of steps; (2) an evaluation of the taxonomy's usefulness through open-ended questions; (3) an evaluation of the instrument's interpretability by means of a questionnaire; (4) an evaluation of the instrument's usability using the System Usability Scale (SUS) (Brooke, 1996); and (5) the opinion of instructors about the instrument after using it.

The structure of this paper is as follows. First, we describe the running case used to generate the taxonomy and the instrument. Second, we detail the taxonomy of activities and questions generated to discover undesired patterns. Third, we explain the development of the instrument and how it answers questions generated before. Fourth, we describe the current methods and tools used by instructors to teach the sequence of steps, evaluate the usefulness of the taxonomy, and evaluate interpretability and usability of the instrument. Finally, we present our conclusions and future work.

4.2. Running Case: The Ultrasound-Guided Internal Jugular Central Venous Catheter Placement

We used the Ultrasound-Guided Internal Jugular Central Venous Catheter (UGIJCVC) placement procedure as a running case to illustrate our approach. UGIJCVC consists of the installation of a tube in the central vein, to assist in the delivery of fluids or medications to a patient. This procedure has the following steps:

(i) Prepare implements, set up the ultrasound equipment, and position the patient.
(ii) Identify the target vein with the ultrasound, and then puncture it with a trocar. A trocar is a needle with a hole to insert the guidewire.
(iii) Verify blood return using a syringe. If it happens, the trocar was correctly installed. Then, remove the syringe.
(iv) Pass a guidewire through the trocar. Once the guidewire is in the vein, remove the trocar.
(v) Widen the pathway and insert the catheter using the installed guidewire.
(vi) Remove the guidewire and install the catheter.

Essential materials to do a process-oriented analysis of the UGIJCVC case were a process model (modeled using BPMN notation) and an event log with executions of the process (Munoz-Gama et al., 2019).

To generate the process model, an initial model was generated using activities included in validated checklists. To avoid a biased mode, de la Fuente et al. (2018) conducted a Delphi panel with experts on the procedure through an online survey. They included the activities in the model when the panel reached 80% consensus for each activity, thus obtaining the ideal execution of the procedure.

Data used correspond to executions performed by ten residents of a simulation-based training course at Pontificia Universidad Católica de Chile (Corvetto et al., 2017), where students received training on the UGIJCVC placement, and students enrolled in the course were given process-oriented feedback (Lira et al., 2019). We obtained data using a web-based software called POMElog (Leiva et al., 2019). This software allowed us to generate event logs from videos recorded while students performed the procedure. Each event of the event log used contains a student as case identifier, a procedure step as activity, and date and time when the student performed the procedure step as the timestamp. An expert manually tagged each video with the activities shown in the process model (see Figure 4.1). Once all videos were tagged, we obtained the event log used in this article.

This study has the approval of the Pontificia Universidad Católica de Chile ethics research committee (ID: 16-194).

4.3. Taxonomy of Activities

The first contribution of this paper is a taxonomy for a procedural skills training course. Instructors need to understand the performance of their students easily, but procedures commonly have many steps, and the sequence of steps followed by each student is different. Therefore, information about how students are learning the sequence of steps could be challenging to understand for instructors. With a taxonomy of activities, it is possible to label the steps with a specific category, produce more synthesized information, and analyze students' performance more easily.

We generated a taxonomy using the semantic of the BPMN model provided by Munoz-Gama et al. (2019). This semantic is related to the domain where the model is applied, represents what surgeons think when they perform surgical procedures, and allows us to classify the activities (i.e., steps) of this model in the following four categories:

1. **Preparation** activities.
   Steps previous to the beginning of the procedure. These steps correspond to the preparation of the patient and the implements needed for the execution of the surgical procedure.

2. **Identification** activities.
   Steps to recognize and locate a structure (e.g., vein and lung) that will be intervened during an action activity. The execution of these steps is always before an action activity.

3. **Action** activities.
   Main steps of a surgical procedure. They represent steps that indicate progress along the stages of the procedure.

4. **Control** activities.
   Steps to verify the correct execution of an action activity or to check if the objective of the action activity was accomplished. Thus, they define if it is possible to continue with the next step or they should go back. They are always performed after an action activity.

5. **Other** activities.
   Steps not performed in a simulation context or steps that make no sense categorizing in one of the aforementioned four categories.

We classified activities depending on the task performed: activities that help to know what is needed before performing a procedure (preparation activities), locate the structure (identification activities), execute a main step of the procedure (action activity), and check if the step was done correctly (control activities).

This definition considers that each activity can belong to only one category. In case it is not clear what activity class a step belongs to, it is possible to split it into more steps, and then classify them in any of the proposed categories.

Figure 4.1 shows the BPMN model with the proposed taxonomy applied. The first 11 steps are preparation activities, including hand-washing and patient positioning. Then, identification activities such as ‘Doppler identification’ help to determine where to ‘Puncture’, which is an action activity. After, ‘Blood return’ is the control activity to verify if ‘Puncture’ was done correctly (i.e., to check if the trocar is inside the vein). Later, ‘Guidewire install’ and ‘Remove trocar’ are the next action activities to execute, to then be verified by ‘Check wire in long axis’, ‘Check wire in short axis’, and ‘Wire in good position’, all control activities. Finally, ‘Advance catheter’ and ‘Remove guidewire’ are the last action activities, which should be controlled by ‘Check flow and reflow’ and ‘Check catheter position’, both control activities.

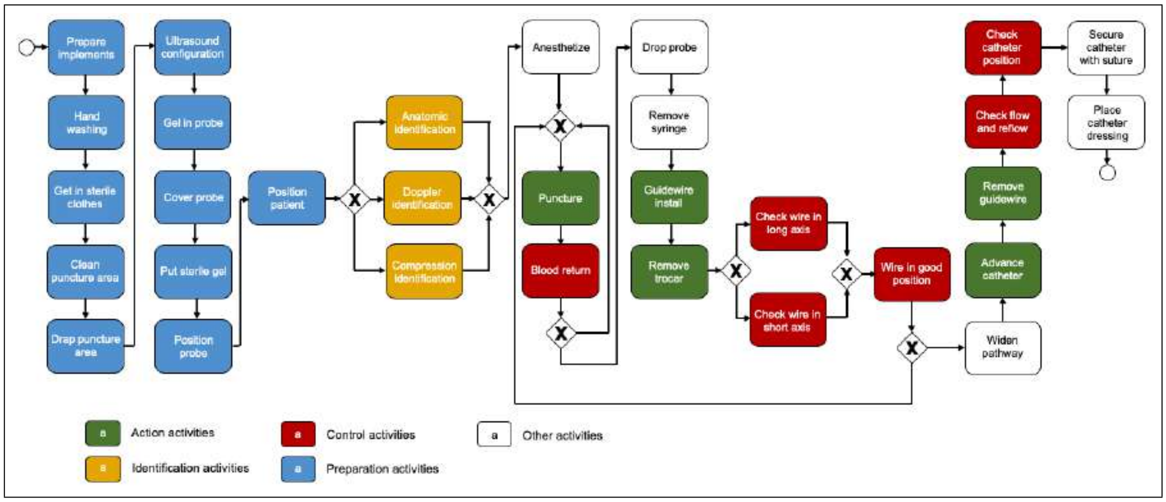


Figure 4.1. BPMN model of the UGIJCVC placement, enriched with the taxonomy proposed.

4.4. Discovering Undesired Patterns

In this section, we present the design of an instrument (see Figure 4.2) to discover undesired patterns. This instrument helps to do a retrospective analysis of the course performance regarding sequence errors. In addition, the instrument provides information at the course level and at the specific student level, allowing the comparison between both levels, knowing the overall performance of the course and thus planning the next sessions, with either current or future students. Creating the instrument involved the following steps: (i) we discovered the process maps depicting residents' performances in Celonis, (ii) we compared the process model with the process discovered in Celonis through visual inspection, to find common mistakes, (iii) after deciding what information would be useful to show residents' mistakes, we designed a first draft of the instrument based on the visual inspection, (iv) we shared the draft with the other researchers in the team and the physician involved to check the instrument's understandability, (v) we iterated between steps (iii) and (iv).

The instrument contains answers for questions designed using the taxonomy presented in Section 4.3. We remark that these questions did not allow us to analyze the successfulness of each isolated step, but helped us to discover undesired patterns related to the sequence of steps. The undesired patterns we looked for were determined based on the process model and the taxonomy.

These patterns emerge from the visual inspection on the process maps discovered in Celonis, and comparing them with the process model. Also, these patterns are undesired from the medical perspective: all the steps have to be performed in the order proposed in the process model, otherwise there are risks for patient safety and the progress in the sequence of steps of the procedure can be compromised. Below, we present the four questions, each with its answer. We also generated the answers using the taxonomy, and then we put them together in the instrument shown in Figure 4.2.

Regarding Process Mining techniques, we used the ideas of algorithms based on Directly-Follows Graphs (van der Aalst, 2019), which are commonly implemented in commercial tools (such as Disco and Celonis). We used this approach because it can be easily understood by non-expert users of Process Mining. Although this approach has some limitations such as a representative bias, it shows the behavior of a process in a simple way (van der Aalst, 2019).

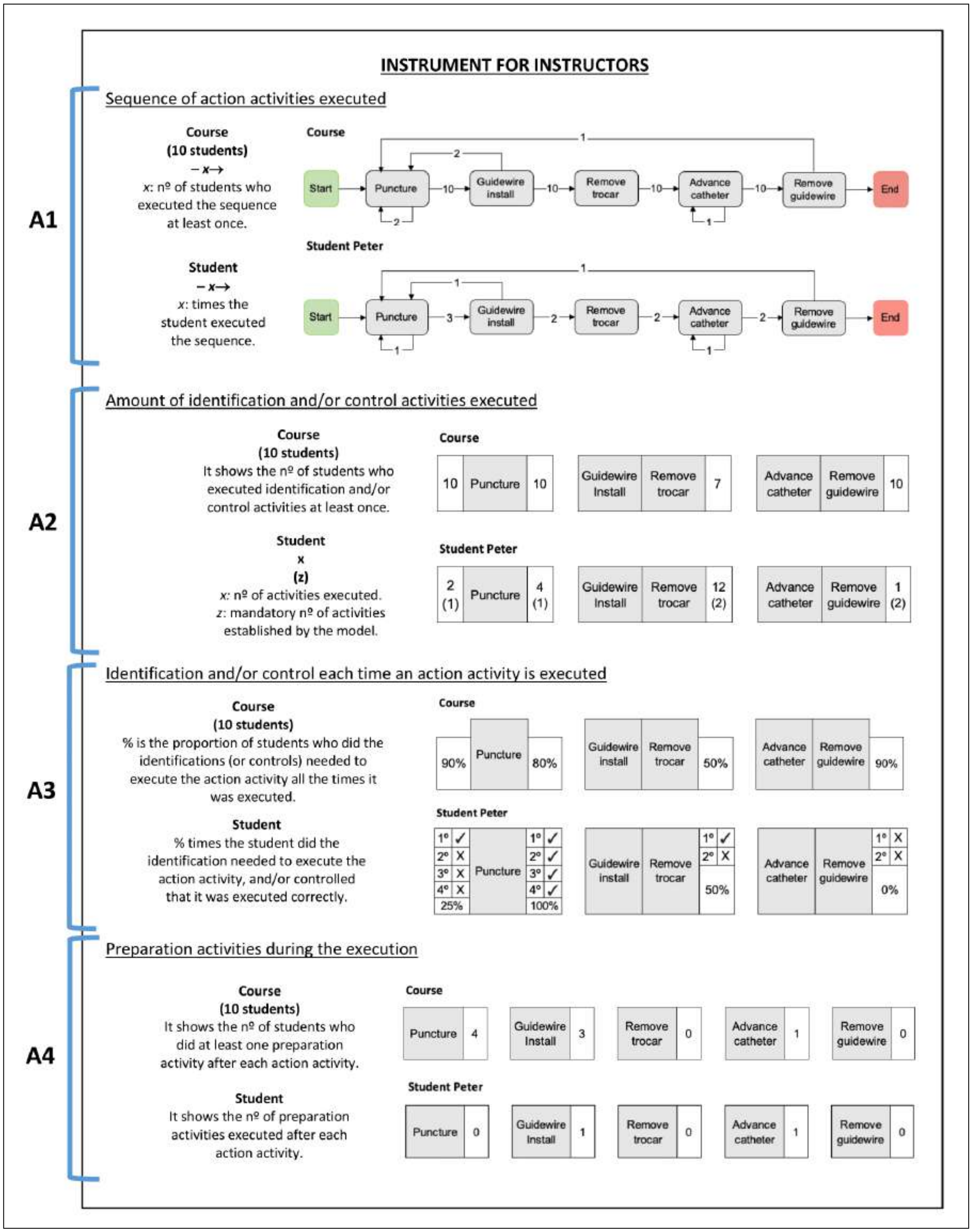


Figure 4.2. Process-oriented instrument for instructors of surgical procedures.

Q1. Which undesired sequence of action steps are students doing?

Execution of surgical procedures consists of following a specific order of action steps correctly. Undesired patterns are going back on action steps, omitting action steps, or repeating unnecessarily an action step.

A1. ***Sequence of action activities executed.***

A1 is the answer for **Q1**. For this answer, we used events in the event log that correspond to *action* activities. It allowed us to know where students went back, did repetitions of a step, or if they omitted any of the main steps of the procedure.

For a student, the answer shows the number of times the student followed the path between two action activities, represented by the arrow. Figure 4.2 (see A1) shows the sequence executed by the student Peter, where is possible to view this student repeated once 'Puncture' and 'Advance catheter', went back once from 'Guidewire install' to 'Puncture', and also went back once from 'Remove guidewire' to 'Puncture'. This student went from 'Puncture' to 'Guidewire install' three times, from 'Guidewire install' to 'Remove trocar' two times, from 'Remove trocar' to 'Advance catheter', and from 'Advance catheter' to 'Remove guidewire'.

For the course, the answer shows the number of students who followed the path between two action activities at least once. In Figure 4.2 (see A1), the view for the course shows that two students repeated 'Puncture' at least once, and one student repeated 'Advance catheter' at least once; two students went back from 'Guidewire install' to 'Puncture' at least once, and one student from 'Remove guidewire' to 'Puncture' at least once. The whole course (ten students) went from 'Puncture' to 'Guidewire install', from 'Guidewire install' to 'Remove trocar', from 'Remove trocar' to 'Advance catheter', and from 'Advance catheter' to 'Remove guidewire' at least once.

Q2. How many identification and control steps were executed?

Some steps of surgical procedures involve the intervention of an organ or part of the body. Before intervening, it is essential to identify the organ, i.e., locate the structure that will be intervened. Other steps check the installation of an instrument or verify if an action had the expected result (e.g., positioning the catheter or other instrument). An undesired pattern is the excessive execution of identification and control steps, because this indicates a lack of fluidity and economy of movement (I. W. Y. Ma et al., 2012).

A2. ***Amount of identification and/or control activities executed.***

A2 is the answer for **Q2**. For this answer, we used events in the event log that correspond to *action, identification*, and *control* activities.

It allowed us to know if students are making an excessive or insufficient amount of identification and/or control activities that are related to each action activity.

For a student, the answer shows the number of events that are identification activities (left side) and/or control activities (right side) for each action activity. In addition, in parenthesis is the desired amount of events. If there is nothing on the right or the left side of the action activity, it means it is not necessary to perform the absent side. Figure 4.2 (see A2) shows the number of activities executed by student Peter. The student did two identification activities before 'Puncture' and four control activities, but the desired amount is once. Peter did twelve times control activities of 'Guidewire install' and 'Remove trocar', but the desired amount is two. Besides, Peter did control activities once after 'Advance catheter' and 'Remove guidewire', but the best approach is to perform two control activities.

For the course, the answer shows the number of students who perform at least one identification (left side) and/or control (right side) activities for each action activity. If there is nothing on the right or the left side of the action activity, it means it was not necessary to perform the absent side. Figure 4.2 (see A2) shows that all students performed some activities before and after 'Puncture', seven students executed some activities after 'Guidewire install' and 'Remove trocar', and all students did some activities after 'Advance catheter' and 'Remove guidewire'.

Q3. **Were identifications and controls executed correctly?**

Besides the number of times executed, it is essential to know if students are identifying and controlling an action step correctly each time it is executed. An undesired pattern is not controlling or identifying each time an action step is executed.

A3. ***Identification and/or control each time an action activity is executed.***

A3 is the answer for **Q3**. For this answer, we used events in the event log that correspond to *action, identification*, and *control* activities.

It allowed us to know if students performed the identifications and/or controls required by the BPMN model each time students executed the action activity.

For a student, the answer in Figure 4.2 (see A3) shows the number of times the action activity was executed. The number is accompanied by '✓' when an identification and/or control was correctly done, or 'X' if it was done incorrectly. A3 in Figure 4.2 shows identification (left side) and/or control (right side) for each action activity. If there is nothing on the right or the left side of the action activity, it means it was not necessary to perform the absent side.

To determine the correctness, we created the following rules:

Considering that a procedure can have one or more identification activities performed in different sequence rules (e.g. when exclusive or parallel gateways are in place), we labelled the identification part as correct:

- When the student performed the identification activity just before the action activity involved, if the model defines only one identification activity.
- When the student performed the identification activities in the way defined by the model (i.e. following the rules depicted in the process model), if the model defines more than one identification activity (such as in Figure 4.1), .

Similarly, we labelled the control part as correct:

- When the student performed the control activity just after the action activity involved, if the model defines only one control activity.
- If the model defines more than one control activity (such as in Figure 4.1):
    - When the student performed the control activities in the way defined by the model (i.e. following the rules depicted in the process model).
    - When the student performed only one control activity, and after performed an activity expected to be executed before the action activity involved, according to the process model. Despite the control was not completely aligned with the process model, this behavior shows that the control activity performed was enough for the student to realize the committed mistake.

Figure 4.2 (see A3) shows the results for student Peter. This student did ‘Puncture’ four times, but only performed the identification correctly the first time (1 of 4, 25%), and performed the control of ‘Puncture’ every time (4 of 4, 100%). Peter performed ‘Remove trocar’ two times and only performed the control correctly the first time (1 of 2, 50%). Besides, Peter performed ‘Remove guidewire’ two times, and did not do the control correctly both times.

For the course, this answer shows the percentage of students who performed correctly the identification (left side) and/or control (right side) every time the action activity was executed. In Figure 4.2 (see A3), the view for the whole course shows that 90% of students made the identification of ‘Puncture’ every time it was executed, and 80% of them did the control correctly; 50% of students controlled ‘Remove trocar’ successfully every time it was performed; and 90% of students controlled ‘Remove guidewire’ successfully every time it was executed.

Q4. Are the students doing preparation steps during the execution of the procedure? Where?

Before the execution of any surgical procedure, it is essential to perform previous steps to prepare the patient and the implements needed along with the procedure. An undesired pattern is to do preparation steps once the procedure begins because it indicates bad preparation by students in the performance of the procedure.

A4. ***Preparation activities during the execution.***

A4 is the answer for **Q4**. For this answer, we used events in the event log that correspond to *action* and *preparation* activities.

It allowed us to know if preparation activities were executed after an action activity. An ideal execution performs all the preparation activities at the beginning (as is stipulated in the BPMN model, see Figure 4.1). An undesired execution shows preparation activities executed between action activities, and it indicates that the student did not prepare the procedure or the patient correctly.

For a student, the answer shows the number of events that correspond to the preparation activities executed after each action activity. Figure 4.2 (see A4) shows the results for student

Peter and indicate that the student did preparation activities after ‘Guidewire install’ and ‘Advance catheter’ once.

For the course, the answer shows the number of students who performed preparation activities at least once after each action activity. In Figure 4.2 (see A4), the view for the whole course shows four students did preparation activities after ‘Puncture’ at least once, three students after ‘Guidewire install’, and one student after ‘Advance catheter’.

4.5. Evaluating the Taxonomy and Instruments with Instructors

We conducted the evaluation through the following stages shown in Table 4.1. The focus of the first stage was to know the current ways instructors use to teach the sequence of steps, give feedback about the sequence, and assess whether students learned the sequence. Then, we evaluated the ease of understanding and usefulness of taxonomy for typical instructor tasks, and the interpretability and usability of the instrument as well as the opinion of instructors after using the instrument with the instructors.

Table 4.1. Evaluation stages.

Stage	Task
1	Current teaching of the sequence of steps.
2	Taxonomy explanation to instructors and its usefulness.
3	Instrument explanation to instructors.
4	Instrument interpretability by instructors.
5	Usability analysis of the instrument.
6	Instructors’ opinion after using the instrument.

To evaluate the taxonomy and the instrument, we conducted semi-structured interviews to three experts who commonly teach UGIJCVC placement, with the aim of validating through expert opinion (Wieringa, 2014) that both artifacts are useful to detect sequence of steps mistakes. They teach in two institutions in anesthesiology and internal medicine specialties, have 5.7 years of experience on average as UGIJCVC instructors, and 12.3 years of experience performing the procedure on average. Evaluation results were analyzed using qualitative content analysis (Graneheim & Lundman, 2004). We asked instructors to answer questionnaires written on paper, and open-ended questions were recorded. Then, we transcribed the audio and analyzed them, to then create five emerging categories (which are tools, methodology, objectives, content, training structure) coding quotes to interpret the answers. We included participant quotes in the paper, and we identified them anonymously with I1, I2, and I3.

4.5.1. Current Teaching of the Sequence of Steps

We asked instructors to describe how they teach the sequence of steps, give feedback to students, and how they assess the sequence of steps. Questions answered by instructors are in Table 4.2.

Table 4.2. Current teaching of the sequence of steps.

No.	Question
1	How do you teach the sequence of steps in the procedure currently? Do you use any tools or instruments?
2	How do you give feedback to students on their flow during the procedure currently? (For example, about what the next step to be performed is, if any are skipped, if it stops, etc.) Do you use any tools or instruments?
3	How do you assess students' flow and the sequence of steps during the procedure currently? (For example, if the correct sequence of steps is executed, if any are skipped, if it stops, etc.) Do you use any tools or instruments?

Answers to Question 1 of Table 4.2, regarding methods of teaching, varied between all the participants' instructors. One instructor said "I give myself as an example of how the process is carried out" (I1). Two instructors said they teach the sequence partitioning the procedure in stages. In addition, one instructor encourages students to verbalize what they are doing for two reasons: one is "the assistant (...) needs to know what the is doctor doing" (I3), and second "we want the student not only to learn the technique but also to lead the procedure" (I3). Regarding tools, two instructors used a checklist of the procedure, based on the literature. Instructors used it to indicate the steps using the order predefined by the checklist. One instructor said she teaches "mentioning the steps a bit with the checklist in mind but not with the paper in hand" (I2). Another instructor sends students videos of how to perform the procedure and documents with anatomic information weekly prior to training sessions.

Answers to Question 2 of Table 4.2, regarding giving feedback methods, noted that instructors do it without a specific structure or pattern. It is a problem for feedback effectiveness because the reliability and credibility of the feedback is an issue (Sharma et al., 2015). One instructor said "I feel that we are always weak in the feedback" (I1) because "we don't have an objective way to correct mistakes" (I1). Regarding tools, two instructors said they give feedback using a global scale (which is a qualitative assessment and is useful for any procedure), a checklist or by advising students based on their experience performing the procedure. One instructor mentioned "a checklist is quite extensive but it is super meticulous for detail" (I3), and another instructor mentioned that she prefers global scales instead of a checklist because "checklists did not discriminate between experts and novices" (I2). The instructor who considers giving feedback challenging did not mention tools such as checklists or global scales, but said "we try to correct at the time the mistake is made" (I1) pointing out the mistake to the resident immediately.

Answers to Question 3 of Table 4.2, regarding methods of assessment, showed issues mentioned by instructors: "I think we do it in a super qualitative way" (I2), "there is no objective pattern" (I1), "We do not have full standardization between teachers" (I2), and "the checklist is long, it is very extensive, which makes it a bit difficult when you evaluate" (I3). Regarding tools, two instructors use the checklist and the global scale based on Martin et al. (2005), but without putting the focus on the sequence of steps. One instructor assesses subjectively if the sequence executed was correct. Concerning the checklist, one instructor said "(the students were) very clear in saying they preferred a very detailed checklist step-by-step, because it was useful for them later in the formative part" (I3), and another mentioned that "today we do not have a tool or an instrument to evaluate that" (I2).

4.5.2. Taxonomy

We evaluated the taxonomy by asking instructors about dimensions shown on Table 4.3 using a five-point visual analogue scale, where one means 'totally disagree' and five 'totally agree'. In addition, we asked instructors their opinion about the usefulness of the taxonomy for giving feedback and assessment tasks.

Table 4.3. Sentences to rate the taxonomy.

No.	Sentence
1	The taxonomy is easy to understand.
2	The taxonomy facilitates the development of a mental model of the procedure.
3	The taxonomy is applicable to other medical procedures (other than UGIJCVC).

The results show that all the instructors agreed with Sentences 1 and 2 (see Table 4.3). In addition, two instructors totally agreed and one agreed with Sentence 3 (see Table 4.3). All the instructors think taxonomy can help them give feedback and assess procedural skills.

Comments of instructors regarding the taxonomy are “I try to think about how to improve or give it another classification and I can’t” (I2), “I think it is really good, the structure helps to carry out a more objective assessment” (I2), “I think that, this process model will be really useful in self-taught training in the future” (I2) and with this, one can say “hey, look you failed on this” or “you did well on this” and “the steps to follow are these” (I3).

4.5.3. Instrument Interpretability

After explaining the instrument to instructors, we asked them to answer an interpretability test, and thus we evaluated their understanding of the information given by the instrument. It consists of asking instructors questions about each answer given by the instrument. The instrument (see Figure 4.2) was generated using a real student and the course in (Munoz-Gama et al., 2019). Interpretability test questions are in Table 4.4.

Table 4.4. Interpretability test questions.

Answer	No.	Question
A1	1.1	What activities did the course repeat?
	1.2	How many students did 'Puncture' after 'Remove guidewire'?
	1.3	What activities did Peter repeat?
	1.4	How many times did Peter do 'Puncture' after 'Remove guidewire'?
A2	2.1	How many students did the identification previous to 'Puncture'?
	2.2	How many students did the control of 'Guidewire install' and 'Remove trocar'?
	2.3	Did Peter do the ideal number of identification activities prior to 'Puncture'? If you have a negative response, did he do more or less than the ideal number?
	2.4	Did Peter do the ideal number of control do activities after 'Advance catheter' and 'Remove guidewire'? If you have a negative response, did he do more or less than the ideal number?
A3	3.1	What percentage of the course did the necessary identification each time they performed 'Puncture'?
	3.2	What percentage of the course did the necessary control each time they performed 'Advance catheter' and 'Remove guidewire'?
	3.3	Peter performed 'Puncture' 4 times. In which of them did he do the identification? In which of them did he do the control?
A4	4.1	How many students did preparation activities right after 'Puncture'?
	4.2	How many students did preparation activities right after 'Remove Guidewire'?
	4.3	How many activities did the student Ana do right after 'Withdraw Trocar'?

The results show that all instructors answered the majority of questions on Table 4.4 correctly (88.1% correct answers on average by each instructor, standard deviation = 3.37%), confirming the success of this Interactive Pattern Recognition case (Fernández-Llatas et al., 2013) i.e., the involvement of domain experts in the patterns interpretation and discovery to generate valid patterns was successful. All instructors answered all the questions related to the number of identification/control activities executed and procedure preparation correctly (Answers A2 and A4, respectively). However, instructors answered some questions regarding repetitions or reworks of action activities incorrectly (two instructors answered Question 1.1 incorrectly and one instructor answered Question 1.3 incorrectly), as well as questions related to the correct execution of identification and control (one instructor answered Question 3.2 incorrectly and another instructor answered Question 3.3 incorrectly).

4.5.4. Instrument Usability

Usability was defined by Brooke (1996) as the appropriateness of an artifact for a specific purpose. Regarding Point 5 of Table 4.1, we used the System Usability Scale (SUS) to determine usability of our instrument (Brooke, 1996), a widely accepted questionnaire to evaluate it. SUS questions are in Table 4.5. Instructors answered each question using a five-point visual analogue scale, where one means 'strongly disagree' and five 'strongly agree'.

Table 4.5. System Usability Scale (SUS) to evaluate the usability of the instrument.

No.	Question
1	I think that I would like to use this instrument frequently.
2	I found the instrument unnecessarily complex.
3	I thought the instrument was easy to use.
4	I think that I would need the support of an expert to use this instrument.
5	I found the various components of this instrument were well integrated.
6	I thought there was too much inconsistency in this instrument.
7	I would imagine that most people would learn to use this instrument very quickly.
8	I found the instrument very difficult to use.
9	I felt very confident using the instrument.
10	I needed to learn many things before I used this instrument.

The mean score was 89.2 (SD = 9.2). According to Bangor, Kortum, and Miller (2008), it means our instrument has an acceptable level of usability. Analyzing the questions, it is possible to see that all instructors agreed or strongly agreed about the frequency with which they would use the instrument, ease of use, integration of the components, fastness of learning to use the instrument, and confidence using the instrument (Questions 1, 3, 5, 7 and 9 in Table 4.5). All instructors disagreed or strongly disagreed regarding the complexity of the instrument, inconsistency in the instrument and difficulties using the instrument (Questions 2, 6, and 8 in Table 4.5). One instructor agreed with the need for expert support to use the instrument while two instructors disagreed or strongly disagreed with this sentence (Question 4 on Table 4.5). The question related to the need for learning things before using the instrument had the same trend (Question 10 on Table 4.5).

4.5.5. Instructors Opinion after Using the Instrument

To describe the opinion of instructors about using the instrument, we asked them open-ended questions after they used it to answer the interpretability test (see questions in Table 4.6). In addition, we asked instructors about the usefulness of the instrument giving feedback and assessment tasks.

Table 4.6. Questions to get instructors opinion about using the instrument.

No.	Question
1	Do you think the instrument helps you know if the students know the correct sequence of steps in the procedure? Explain briefly.
2	What is your opinion of this instrument, compared to the tools/instruments that you commonly use? Explain briefly.
3	What was the most difficult thing about using the instrument?
4	What would you improve about the instrument?

Regarding Question 1 of Table 4.6, instructors said the instrument helps to do a more objective assessment, because even “one omit steps when performing the procedure, and therefore also skips steps when assessing a resident” (I1), which is an issue supported by the literature (Sullivan et al., 2014). In addition, one instructor said “this can help us know mentally if the student knows the next step, both the instructor and the student” (I2) and another instructor mentioned the “different items of the instrument served to understand that sequence” (I3).

Answering Question 2 of Table 4.6, instructors highlighted the characteristics of the instrument: one instructor mentioned “the comparison with the course is always more attractive than the individual as a single entity” (I3). In addition, they mentioned the possibility of tasks they can do with the instrument in comparison with others: one instructor said “one can compare with previous years, with other groups and with other groups from other institutions” (I1), “it is often challenging to train the person to use the global scale, I think this will be easier to train in because it is very logical” (I2), and “I would imagine that showing this to a student would make it easier for him/her to understand why it is important and why I am giving him/her this feedback or mark” (I3). Additionally, the instrument allowed instructors to reflect on their performance as a teacher: one instructor said “(the instrument) lets me know as an instructor if there is a step that I am not explaining well or there is something that needs more reinforcement” (I3), “I think it is exciting to know where I am strong and where I need to put more emphasis on preparation on the course level as well as on the individual level" (I3).

Answering Question 3 of Table 4.6, two instructors said the first component was the hardest to understand, but, once we explained it to them, the items were easy to understand. Regarding Question 4 of Table 4.6, a suggestion was to create material to facilitate the instrument interpretability (e.g., a video with an explanation of how to interpret it).

4.6. Discussion

In this paper, we present a taxonomy of steps for surgical procedures and an instrument for instructors, both focused on teaching the sequence of steps required to perform a surgical procedure. Both tools help to obtain process-oriented information that could be useful for surgical instructors during training. The taxonomy developed can be used by instructors in everyday tasks, such as giving feedback and assessing students. Instructors can do the same tasks using the instrument, which shows information related to mistakes in the sequence of steps through the components it provides.

Instructors agreed that the taxonomy could help to give a structure to training, and they found it useful for giving feedback and assessing students' performance. Regarding the instrument, instructors understood the information provided and considered the usability of the instrument acceptable. Similar to taxonomy, instructors thought the instrument helped them establish if students knew the correct sequence of steps, and it could be useful for everyday tasks they do as surgical teachers. Hence, we conclude that the information provided by the instrument (generated using the taxonomy) could be useful to understand students' performance as regards the sequence of steps. Therefore, the taxonomy and the instrument are resources to include in the tools instructors have to teach surgical procedures.

The taxonomy and the instrument can help instructors in building or enhancing their mental model. For instructors, it is challenging to share their mental model (Crebbin et al., 2013; Lateef, 2018), and that means that they omit close to 70% of the information needed by students during their learning process (Sullivan et al., 2014). The taxonomy and the instrument help to address these problems. With the taxonomy, it is possible to share information about the steps of the procedure and their correct sequence; with the instrument, the information utilized to teach the correct sequence is standardized, avoiding the omission of information. In addition, both tools help to give a structure as to how to teach the procedure, providing standardization of the contents and an objective way for assessment and giving feedback.

A limitation of this research is the number of instructors who participated in the evaluation. This issue impacts on the opportunity of having more feedback about improvements for the instrument, for instance, on how to present the information in an easier-to-understand way. Also, the low number of participants makes difficult to identify with statistical significance the sections of the instrument that are more difficult to interpret for instructors. However, we asked instructors from different specialties and institutions; thus, we believe the number of instructors provides sufficient evidence to accomplish the objective of this research. Another limitation is the possibility of difficulties to handle the instrument if the number of steps of a procedure is more extensive than UGIJCVC placement. In that case, figures of answers generated (see Figure 4.2) could change a bit to provide the same information as could be needed. Furthermore, the instrument does not contain subjective information as provided by GRS, because the instrument shows objective information. Such a need could be addressed using the instrument as a complement to checklists and GRS, allowing the instructor to capture subjective and objective information.

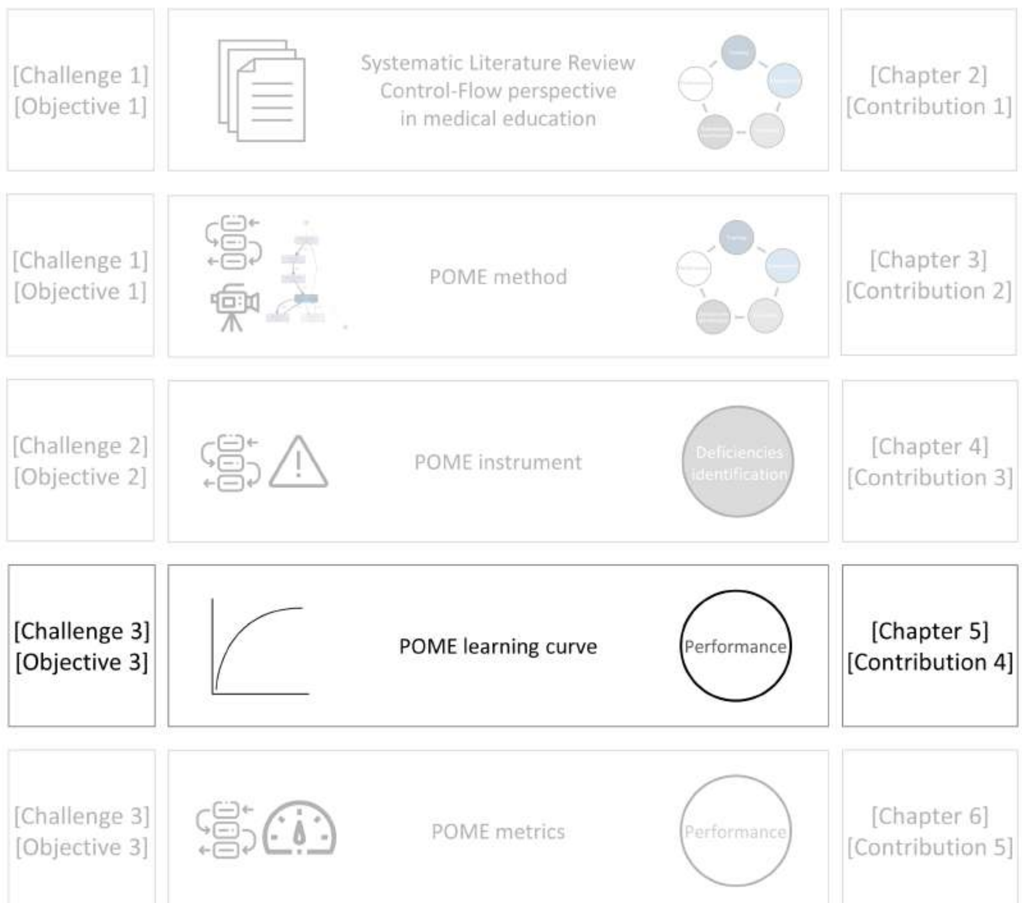
Future work to enhance the instrument's interpretability is an improvement on showing the information in an easier-to-interpret way, which can be inferred from the results obtained by the instructors in the interpretability test. A suggestion to implement such improvement is to provide the same information currently presented in the instrument using natural language (i.e. text in prose) instead of figures that can be confusing for instructors. Another improvement needed is the incorporation of a classification between minor and major sequence errors, ensuring an easy-to-interpret display of this information. In addition, further research is needed to demonstrate whether the taxonomy and the instrument improve the learning of a surgical procedure (here, we did an evaluation using expert opinion (Wieringa, 2014)), to demonstrate whether the instrument helps to improve feedback and assessment of students, and to make the generation of the instrument by potential users available and customizable.

4.7. Conclusions

We present a taxonomy of activities for surgical procedures and an instrument for instructors showing undesired sequence patterns, generated using Process Mining. After evaluation of both tools with experts, we found them as easy to understand, interpretable by instructors, and with an acceptable level of usability. Further studies should be done once instructors gain experience using the tool.

Chapter 5

Percutaneous Dilatational Tracheostomy Training from Control-Flow Perspective: The parts do not equal the whole



5. PERCUTANEOUS DILATATIONAL TRACHEOSTOMY TRAINING FROM CONTROL-FLOW PERSPECTIVE: THE PARTS DO NOT EQUAL THE WHOLE

5.1. Introduction

Percutaneous dilatational tracheostomy (PDT) is a frequent procedure performed in the intensive care units (Abe et al., 2018; Singh & Sing, 2019). Multiple reports have shown that there has been a steady increase in its use, and novel techniques have been developed to enhance procedural success, such as ultrasound and bronchoscopic guidance (Rudas et al., 2014; Abbott et al., 2021). PDT is not a risk-free procedure, as 3-7% of patients present minor or major complications, with some being potentially fatal (Díaz-Regañón et al., 2008; Dempsey et al., 2010; D. Young et al., 2013).

Simulation-based training has demonstrated its positive impact in procedural competency acquisition, and even translated into better clinical results (Barsuk et al., 2009; I. W. Y. Ma et al., 2011; Zendejas et al., 2011; Seam et al., 2019). Recently, a simulation-based PDT training program has been described in the literature, in which the PDT learning curve plateaus after 6 executions (Kattan et al., 2020). However, in this report and in others, a negative performance gap persists between simulation-based executions and more complex models in the clinical scenario (Sawyer et al., 2015; Kattan et al., 2020). While a proportion of the gap can be explained by differences between the simulated scenario and a real clinical situation, another proposed factor is that current metrics and analyses do not reliably detect when learning is complete (Stefanidis et al., 2012).

The incorporation of new technologies in the operating room has made it possible to generate new information for the analysis of procedural skills with objective metrics, e.g., intraoperative use of energy devices (Hosogi et al., 2021) and type and number of dissection movements (R. Ma et al., 2021). However, PDT and a large percentage of bedside procedures are not technologically supported, to obtain objective data. Considering procedures as surgical processes, i.e., as a set of one or more linked activities whose instances are intended to collectively realize surgical objectives, would allow a quantitative analysis based on data using Process Mining techniques. Process Mining is an emerging discipline that bridges the gap between data science (data mining, machine learning, etc) and model-based process analysis (process modeling, business process management techniques, etc) (van der Aalst et al., 2012; van der Aalst, 2016). Process mining approach allows, based on control-flow perspective and through mathematical tools, comparisons between normative models of a process and real executions (Munoz-Gama, 2016; Carmona et al., 2018). This type of approach would make it possible to obtain data on how a procedure is executed during the evolution of a training and to measure its similarity to a normative model of the procedure, through the objective quantification of omissions, repetitions or alterations in the sequence of activities of the procedure.

Learning curves are intended to describe the residents' performance across a training course with a certain number of sessions. However, as previously stated, it is possible that the objective metric to build the existing learning curves are not capturing an aspect that can explain the gap between performing the procedure in a simulator vs in a real clinical scenario. Therefore, we propose the design of a learning curve based on an objective metric that captures the PDT control-flow aspect, to explain the named performance gap. After finding explanations for this gap, it will be possible to detect the phases in which residents are committing errors, allowing residents to transition from the novice to the advanced beginner competence level (Khan & Ramachandran, 2012).

Our objective was to use Process Mining analysis to obtain novel information about the students' learning progress in a training course. We hypothesized that with the new information, we can build a novel procedural learning curve that will deliver new insights of PDT procedural learning. This findings will tackle the paucity of research and lack of agreement in these aspects (Massick et al., 2000; Petiot et al., 2017; Raimondi et al., 2017), and also contribute to ensure adequate training and deliver competent patient care..

5.2. Materials and Methods

5.2.1. Ethical approval

The Ethics Committee of the Faculty of Medicine, Pontificia Universidad Católica de Chile, approved the original study (reference number 180704005).

5.2.2. Training program description

This study was a secondary analysis of a previously published training program of PDT (Kattan et al., 2020). In brief, the training protocol was performed at the Simulation Center of the Pontificia Universidad Católica de Chile, between January 2019 and July 2019. The cohort was composed of four anesthesiology residents, two emergency medicine residents and two intensive care fellows. The training protocol was based on the mastery learning framework (McGaghie, 2015), and consisted of at least six training sessions in a low-cost simulator (Kattan et al., 2019) plus a final session in a cadaveric model to measure skill transferability (Kattan et al., 2020). Personalized feedback was given after each execution, with the aid of a procedural flowchart (de la Fuente et al., 2021).

5.2.3. PDT surgical process model

In this study we used a surgical process representation of PDT execution. This representation is based on a comprehensive model previously developed through a Delphi panel with Spanish-speaking international experts (de la Fuente et al., 2021). The original process model provides different options to perform the procedure depending on the available resources, and has 59 steps. We used a simplified variant that has 23 relevant steps (see Figure 5.1) to the simulation context of the study.

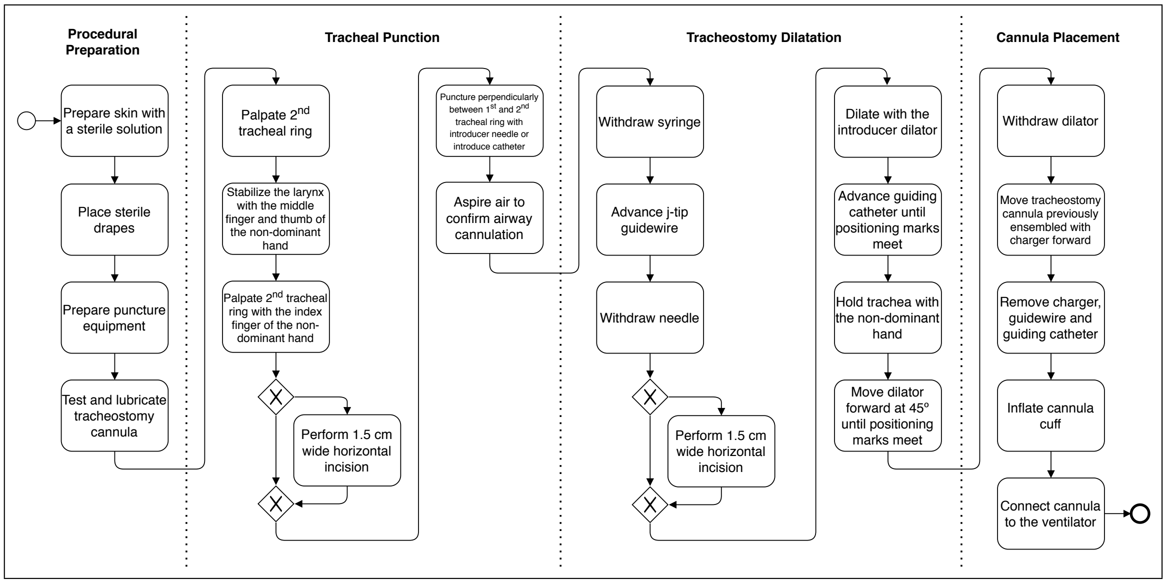


Figure 5.1. Process model with the PDT steps and their order. The model contains four sequences of steps: the one with "Perform 1.5 cm wide horizontal incision" step in the "Tracheal Punction" and "Tracheostomy Dilation" stages; another one with the same step only in the "Tracheal Punction" stage; another with the same step only in the "Tracheostomy Dilation" stage and another that does not contain the step in either of the two stages. We used the sequence that produced the highest similarity.

5.2.4. Data collection

To perform this study, we needed to determine the specific steps and the order in which each resident performed them. We retrieved recorded videos of each resident performing the complete procedure in each session, thus having six videos performing the procedure in a simulator and one in cadaver for each resident. We uploaded the videos to the POMElog platform (Leiva et al., 2019) (adequately encrypted to ensure data safety), where the videos were tagged using the procedure's steps in Figure 5.1 as tags. The videos were tagged by an observer blinded to the performer and session number.

5.2.5. Similarity metric

Using the data about the order of steps performed by the residents, we calculated the similarity between the residents' order of steps versus the order defined by the process model in Figure 5.1.

The similarity metric is based on the Normalized Levenshtein Distance (NLD), that compares two character sequences on the basis of counts the character insertions, deletions and substitutions needed to make both sequences of characters equal, and then divides this amount by the sum of the length of both strings (Yujian & Bo, 2007). We calculated the similarity as one minus the NLD value, so the metric varies between zero and one (see example on Figure 5.2). The interpretation of the metric is as follows: if the similarity is zero, the order of steps performed by the resident was completely different from the order defined by the process model; on the other hand, if the similarity is one, the order of steps performed by the residents was equal to the order defined by the process model. To use NLD as a similarity metric, each activity defined in the normative model was assigned a specific character, which allowed each run to be represented as a sequence of characters and thus compared to the sequence of characters represented by the normative model. Since the process model allows omitting some steps, we implemented the following algorithm to obtain the similarity metric for each resident's performance: (i) we obtained all the traces accepted by the process model (in this case, four model traces); (ii) we calculated the similarity metric comparing the resident performance with each model trace obtained; (iii) we selected the trace model with the highest similarity to calculate the similarity metric for each resident.

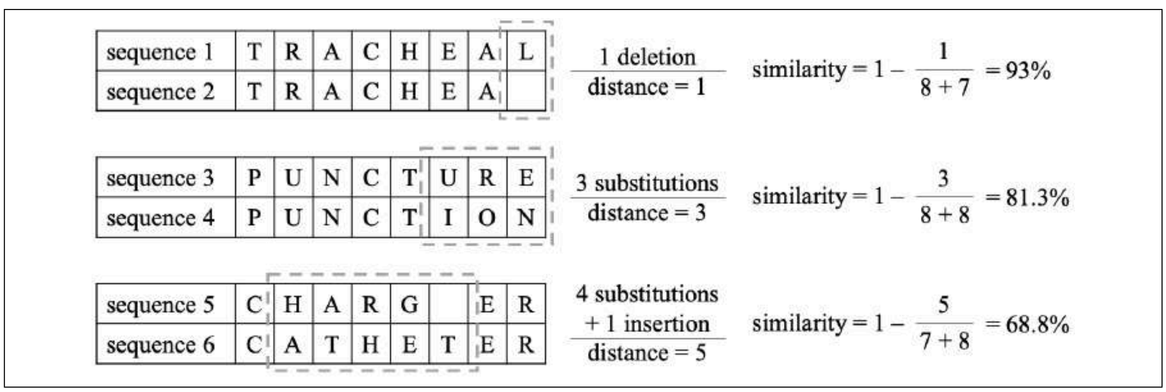


Figure 5.2. Example of similarity calculation. Let's suppose that the words are composed of a sequence of letters and we want to calculate the similarity between both sequences. When comparing sequence 1 (length 8) vs sequence 2 (length 7), we see that the letter L is deleted, so the distance between both sequences is 1. When comparing sequence 3 (length 8) vs sequence 4 (length 8), we see that 3 letters were substituted (U by I, R by O, E by N), so the distance is 3. When comparing sequence 5 (length 7) vs sequence 6 (length 8), we see that 4 letters were substituted (H by A, A by T, R by H, and G by E) and that the letter T was added in sequence 6, therefore, the distance is 5.

5.2.6. Statistical analysis

We evaluated whether the difference in similarity between the first and sixth session was statistically significant, for the similarity of the entire procedure and for each stage. We conducted the same evaluation between the sixth and cadaver session. For the similarity of the entire procedure, we also applied the statistical test between consecutive sessions (first versus second session, second versus third session, and so on). We used the two-tailed Wilcoxon signed-rank statistical test considering $P < .05$ as statistically significant, which compares the difference between the similarity metric in each involved session for each resident. If $P < .05$, it means that there are a statistical diference on similarity between both sessions. We performed the statistical tests using the open-source library SciPy 1.4.1 (SciPy Developers). Furthermore, we made the graphs following Pusic et al. recommendations (Pusic et al., 2015) to show the evolution of the median and each resident. We made the graphics with the open-source library Matplotlib 3.2.1 (Matplotlib Development Team).

5.3. Results

The simplified PDT normative model (Fig. 5.1) establishes the execution of the procedure in 22 steps. The median number of steps performed was 20 steps, the longest performance was 29 steps and the shortest was 15 steps, due to repetitions and/or omissions of steps.

5.3.1. Similarity of the entire procedure

Figure 5.3 shows the evolution of similarity as the sessions progressed. The median curve indicates that the similarity of the entire group of residents increased as the sessions progressed and that it remained stable between the fifth, sixth and cadaver sessions. Also, the entire group of residents converged to a more similar performance to the order of steps present in the normative PDT process model, which is explicit in the decrease of the interquartile range's size as the sessions progressed (see boxplot in Figure 5.3).

Table 5.1 presents the similarity median and interquartile range (Q1 - Q3) between the normative PDT process model and the residents' executions in the first, sixth and cadaver sessions. The difference between the similarity of the first and sixth sessions was statistically significant ($P = .03$)

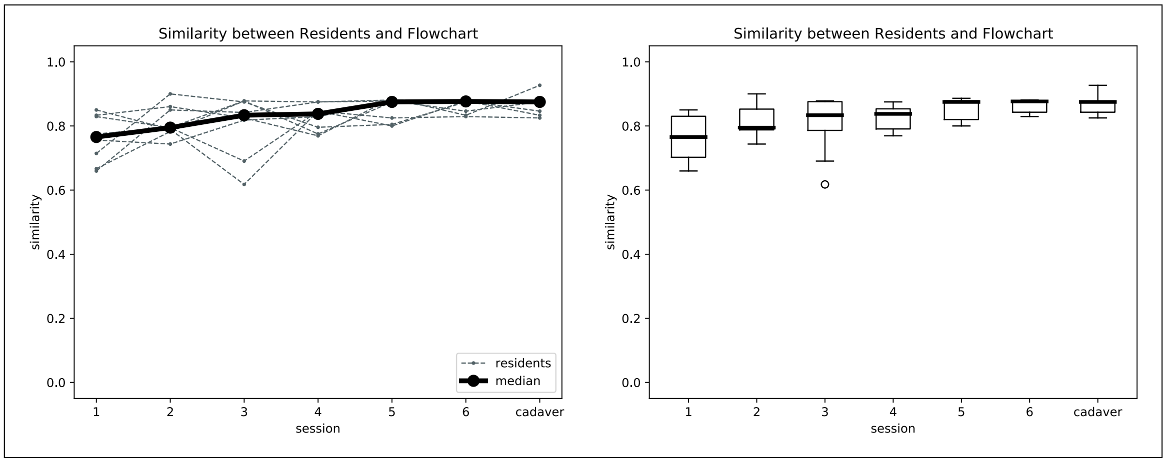


Figure 5.3. Similarity between residents and process model for the entire procedure, from a PDT training cohort at the Simulation Center of the Pontificia Universidad Católica de Chile, 2019. The boxplot's dashed line shows that the difference between the first and sixth sessions is statistically significant.

Table 5.1. Similarity residents versus flowchart for the entire procedure and for each stage, from a BG-PDT training cohort at the Simulation Center of the Pontificia Universidad Católica de Chile, 2019.

Stage	**First session, median (Q1-Q3)**	**Sixth session, median (Q1-Q3)**	**Cadaver session, median (Q1-Q3)**	**p-value first vs sixth sessions**	**p-value sixth vs cadaver sessions**
Entire procedure	0.77 (0.70-0.83)	0.88 (0.84-0.88)	0.88 (0.84-0.88)	**.03**	.75
Procedural Preparation	0.86 (0.75-0.86)	1.00 (1.00-1.00)	1.00 (1.00-1.00)	**.02**	N.A.
Tracheal Punction	0.67 (0.64-0.71)	0.67 (0.65-0.67)	0.67 (0.67-0.79)	.46	.50
Tracheostomy Dilatation	0.73 (0.71-0.81)	0.86 (0.85-0.86)	0.85 (0.85-0.86)	.06	.34
Cannula Placement	0.79 (0.43-1.00)	1.00 (0.95-1.00)	1.00 (0.95-1.00)	.11	1.00

Abbreviations: N.A. means not applicable (all the residents did the preparation well in sixth and cadaver sessions).

5.3.2. Similarity by stages

Figure 5.4 shows the similarity of each stage of the procedure. Table 5.1 presents, for each stage, the similarity median and interquartile range (Q1 - Q3) between the normative PDT process model and the residents' executions in the first, sixth and cadaver sessions. The difference between the similarity of the first and sixth session of Procedural Preparation stage has a statistically significant similarity difference between both sessions (P < .05). Conversely, the similarity difference in Tracheal Puncture, Tracheostomy Dilatation and Cannula Placement stages was not statistically significant (P > .05). None of the similarity differences between the sixth and cadaver session were statistically significant (P > .05).

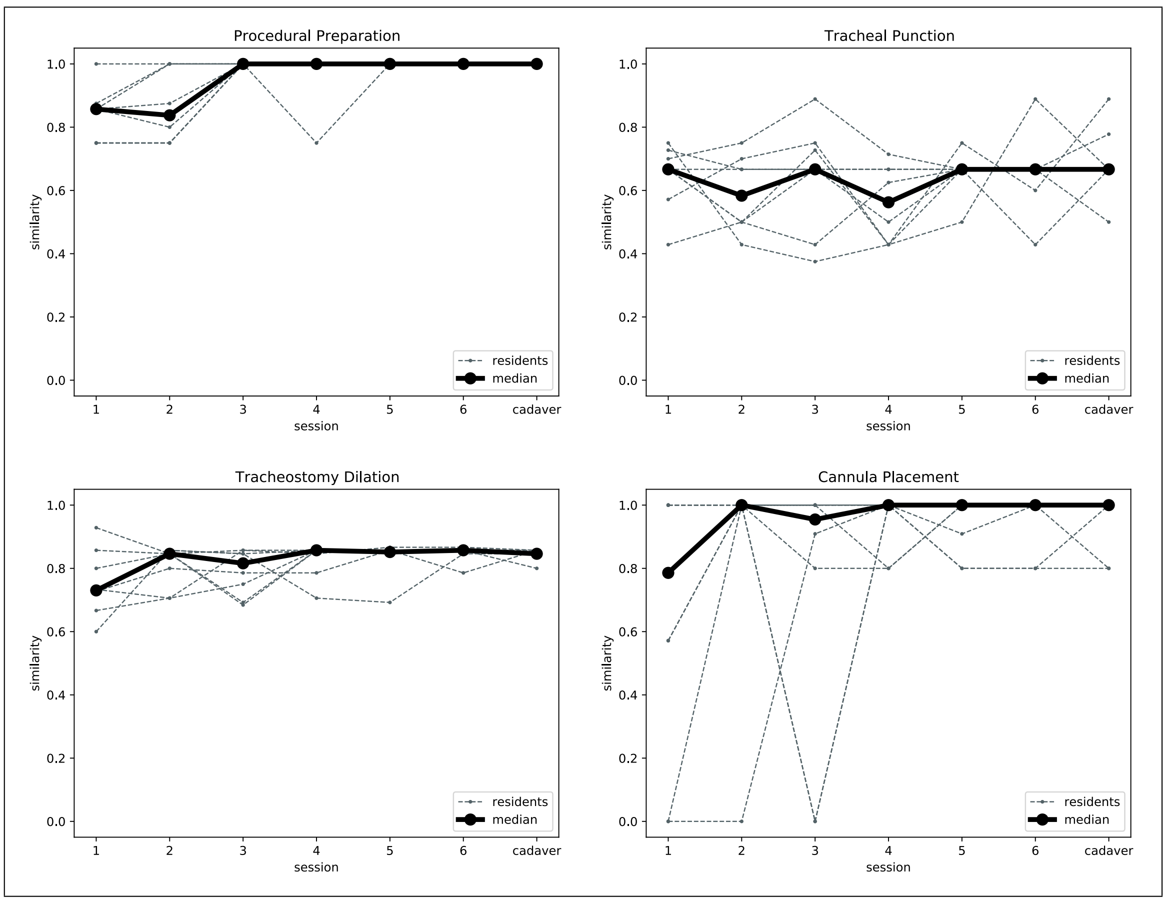


Figure 5.4. Similarity between residents and process model by stage, from a PDT training cohort at the Simulation Center of the Pontificia Universidad Católica de Chile, 2019.

5.4. Discussion

Our study shows that a control-flow analysis of PDT execution videos is feasible and provides novel information about novices' learning progress. The learning curve shows that similarity between students' performance and the normative process model increases through the sessions. However, when we break down the analysis by stage we see that students learned some stages at the end of the course, and that the learning of some stages did not improve throughout the course.

The original training program description study concluded, based on the Global Rating Scale (GRS) residents' scores, that all the participants achieved the BG-PDT mastery criteria (GRS score of 21) in the sixth session (Kattan et al., 2020). However, in our curve the similarity stabilized from the fifth session onwards at the procedure level, and in the last sessions there were still residents who did not perform the order of steps as indicated in the normative process model. Furthermore, our analysis by stage showed that, although some stages were very well learned (such as the preparation and cannula placement stages), tracheostomy puncture stage has the same similarity at the start and at the end.

One explanation for the diverging evolution of the intermediate stages similarity metric is that they involve different cognitive loads for the trainees. Several ways of estimating cognitive load have been proposed: self-report tools (psychometric scales) and non-self-report tools (response time and accuracy in a secondary task, brain imaging, heart rate, skin conductance, among others) (Dias et al., 2018; Sewell et al., 2019). The similarity metric we used combines the sequence errors (steps omitted, repeated and executed in a different sequence) to calculate its value, which can be considered a measure of the cognitive load that residents had at each stage of the task. This type of analysis could explain the rapid increase in similarity metrics of the Procedural Preparation stage. This stage has common activities with a variety of procedures, which are already stored in the residents' long-term memory. This fact facilitates the incorporation of this stage to their long-term memory with the inherent variations of this new procedure through long-term memory storage factors such as elaboration, reflection, organization and meaningfulness (Schunk, 2012). At the other extreme, the Tracheal Puncture stage does not achieve improvements throughout the sessions. This finding could imply that this task results in a working memory overload that fails to be processed throughout the course (Fraser et al., 2015; Sewell et al., 2019), especially considering that the trainees repeated the entire procedure without training in specific stages.

These findings are significant because there are few references regarding how to develop training programs of PDT, allowing us to recommend different strategies to deal with this uneven progress, for example: dividing the training into stages (Segmenting Effects), using videos or simpler models of the stages (Pretraining Effects) and concurrent use of spoken text when executing these steps (Modality Effect) (Fraser et al., 2015). Besides, this new information can help them to know the most difficult stages and thus strengthen them when they teach the procedure, as well as knowing the performance of the entire group of residents (Valsamis et al., 2018). On the residents' side, they can know each stage's complexity and decide which stages to pay more attention to, rehearse and focus their learning, allowing them to optimize their germane load (Sewell et al., 2019).

Performance gap between simulator and real life execution is a complex phenomena, and could be influenced by a myriad of factors, including situational awareness (Schulz et al., 2013), assessment stakes, teacher-learner relationship (Schut et al., 2020), cognitive load (J. Q. Young et al., 2014), anatomical and clinical variability of real patients, among others. Another potential factor that could influence this gap, is the ability of our assessment instruments to capture an adequate performance. In the absence of a gold-standard, the main assessment instruments -checklists and global rating scores- suffer from inherent flaws and criticisms (I. W. Y. Ma et al., 2011). Critical performance errors could be overlooked or de-emphasized by the observer, and dragged to real patient performance. In this sense, our approach gives the opportunity to dissect and "zoom-in" into step-by-step performance from the control-flow perspective, identifying potential mistakes not captured by standard instruments.

There is a growing interest in defining objective metrics to better understand the differences between novices and experts in procedural skills (Vedula et al., 2017; Azari et al., 2019). Differences have been described in terms of dissection patterns for robotic surgery (R. Ma et al., 2021), energy device usage patterns in laparoscopic gastrectomies (Hosogi et al., 2021), tool-motion and eye-gaze data in endoscopic sinus surgeries (Ahmidi et al., 2012). However, PDT and other procedures are not technologically supported, making such an approach more difficult. Similar to the PDT analysis, it has been previously used a process mining approach in order to characterize the training in central venous catheter (CVC) placement and demonstrated that the repetition of activities and deviations from the expected sequence of execution are different between novices and experts concentrating on specific stages of the procedure (de la Fuente et al., 2020). In this context, for PDT and CVC placement, the stages that do not reach the expected levels of similarity indicates that they were difficult to perform for novices, and therefore a high level of similarity on these stages could be markers of an advanced level of competence, which should be explored in the future. This information is important in terms of procedure characterization, training design and as possible new assessment metrics.

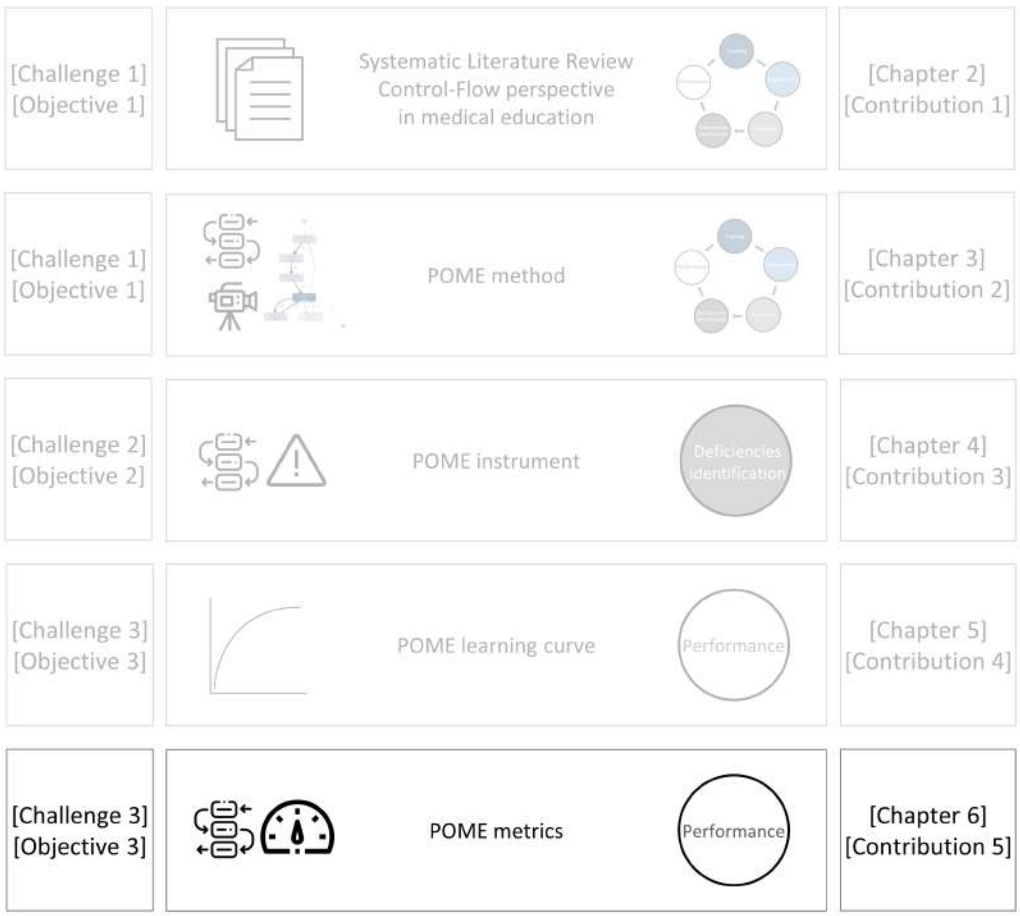
A limitation of our study is the absence of comparison with experts' order of steps. However, the similarity metric that we used is based on a variant of the Levenshtein distance, which differentiates the performance of a novice surgeon from an expert in cutting task, simple suturing and complex suturing in a single and triple incision (Schumann et al., 2013). Also, using the same metric, Schumann, Bühligen, and Neumuth (2015) showed that a performance that follows the order of steps similar to the best practice (in our study, the process model) produces positive surgical outcomes. A second limitation is the resources necessary to obtain the data, since we needed physicians acquainted with PDT in order to label the videos, which is a relevant cost considering their opportunity cost (Roshetsky et al., 2013). Another potential limitation is that in some real life clinical scenarios, due to specific patient or context situations, a strict order of steps can not be readily followed, and adaptations or modifications to the technique can resolve challenging situations. In this sense, the normative model can be further depurated to include the most common ones.

5.5. Conclusions

Residents' performance was increasingly similar to the surgical procedure's best practice process model as they progressed in a simulation-based training course. Similarity by stages analysis shows that some stages increased in similarity at the end of the course, while others end on the same level of similarity as at the beginning of the course. This information helps instructors focus their efforts on strengthening the stages that residents did not learn well and thus optimize their training courses.

Chapter 6

Process-oriented metrics to provide feedback and assess the performance of students who are learning surgical procedures: The Percutaneous Dilatational Tracheostomy case



6. PROCESS-ORIENTED METRICS TO PROVIDE FEEDBACK AND ASSESS THE PERFORMANCE OF STUDENTS WHO ARE LEARNING SURGICAL PROCEDURES: THE PERCUTANEOUS DILATATIONAL TRACHEOSTOMY CASE

6.1. Introduction

Medical education has shown a significant increase in the use of simulation to teach and evaluate procedural skills (Scalese et al., 2008). This training method allows instructors to conduct more learner-centered training (Lammers et al., 2008) and has also been shown to be an effective method for residents to reach an adequate level of proficiency prior to patient contact (Seam et al., 2019). However, its use has been limited among others due to the high costs of simulation models for specific skills (Lichtenberger et al., 2018). There are studies that have sought to reduce their costs for some procedures with new technologies (Lichtenberger et al., 2018; Kattan et al., 2019). Even so, many procedures still remain very expensive to teach and evaluate using simulation. Besides, the opportunity cost of having a doctor teaching surgical procedures to students instead of treating patients is very high (Aitken, 2012). Therefore, it is critical to generate maximum learning for residents from the information obtained in each training session.

Two complementary goals of the instructional process of surgical procedures are to assess performance and provide feedback. The purpose of assessing competency in the performance of a surgical procedure is to define whether a person is capable of performing it under certain conditions. In turn, the objective of feedback is to provide specific information comparing the student's performance and a standard (Ghaderi & Farrell, 2020).

In simulation, several evaluation methods are used to measure the level of competence acquired by residents during training. Commonly, two types of approaches are used in evaluation: Global Rating Scales (GRS) (D. D. Anderson et al., 2016) and Checklists (Epstein, 2007). GRS are subjective, but have the flexibility to be adapted to any surgical procedure. Checklists, on the other hand, seek to demonstrate whether or not each of the steps of the procedure is performed (Lammers et al., 2008). Both have proven to be effective in establishing the level of proficiency in the execution of procedures (Morgan et al., 2001). However, they are designed exclusively for that purpose (Williams et al., 2002). Using checklists do not allow to capture explicit information about the variability with which the resident performs the sequence of steps of each execution nor the evolution of this sequence of steps through the training sessions.

Process Mining is an emerging discipline that generates knowledge from process execution data recorded in information systems (van der Aalst, 2016), facilitating the analysis of the observed process. This discipline has been used in several healthcare specialties (Rojas et al., 2016). Since surgical procedures can be understood as a progression of steps, they can be viewed as a process (Neumuth, 2017), so the inclusion of process mining for their analysis has emerged as an opportunity to deliver new information about learning in these procedures.

Recently, this approach has been applied (de la Fuente, Fuentes, Munoz-Gama, Riquelme, et al., 2020) to analyze the variability with which experts and residents perform the installation of Ultrasound-Guided Internal Jugular Central Venous Catheter (UGIJCVC), identifying patterns that show the difficulty that residents have in learning this procedure. On the other hand, for the same procedure, the use of process mining techniques was proposed to identify desired and undesired process patterns, in order to complement personalized feedback to students using a process perspective (Lira et al., 2019).

In this study, we analyzed a Percutaneous Dilatational Tracheostomy (PDT) training case extracted from a simulation study (Kattan et al., 2020) that, like any common surgical training, is evaluated with metrics that measure the proficiency of the skills of performing the procedure as a whole. Within these, one can find the OSATS (Objective Structured Assessment of Technical Skills, a type of GRS) performance metrics and the execution time, which evaluate the process as a whole. However, the feedback obtained from these metrics may not be sufficient to know where residents may be having problems with the order of steps, so the need arises to look for metrics that complement them and make better use of the available information.

Our hypothesis is that, by using metrics obtained using process mining, we can assess in a more detailed way the proficiency with which residents perform their training sessions and, at the same time, provide more accurate information about where they are making mistakes. Through a top-down analysis focused on stages and activities of the surgical procedure, critical stages and activities in the execution of the procedure are detected, and the variability with which residents perform them, which is not fully captured by classical evaluation metrics, becomes evident.

6.2. Materials and Methods

In this study, the use of process-oriented metrics is proposed to analyze the performance of residents during their training in a PDT procedure simulator. The study is based on the POME method (Munoz-Gama et al., 2021), which is basically composed of three stages: Model stage, Data stage and Analysis stage.

6.2.1. Model Definition

A reference process model was established to extract and analyze the activities of the procedure. For its definition, a generic model of the Bronchoscopy-Guided PDT (BG-PDT) procedure was used, which was developed based on the consensus among experts using the Delphi method (de la Fuente et al., 2021). This BG-PDT model was adapted to represent the procedure incorporating the limitations of the simulator. Finally, a more reduced model was obtained (Figure 6.1), which considers a total of 23 activities (21 mandatory and 2 optional), which are grouped into 4 stages: procedural preparation, tracheal puncture, tracheal dilatation and cannula placement.

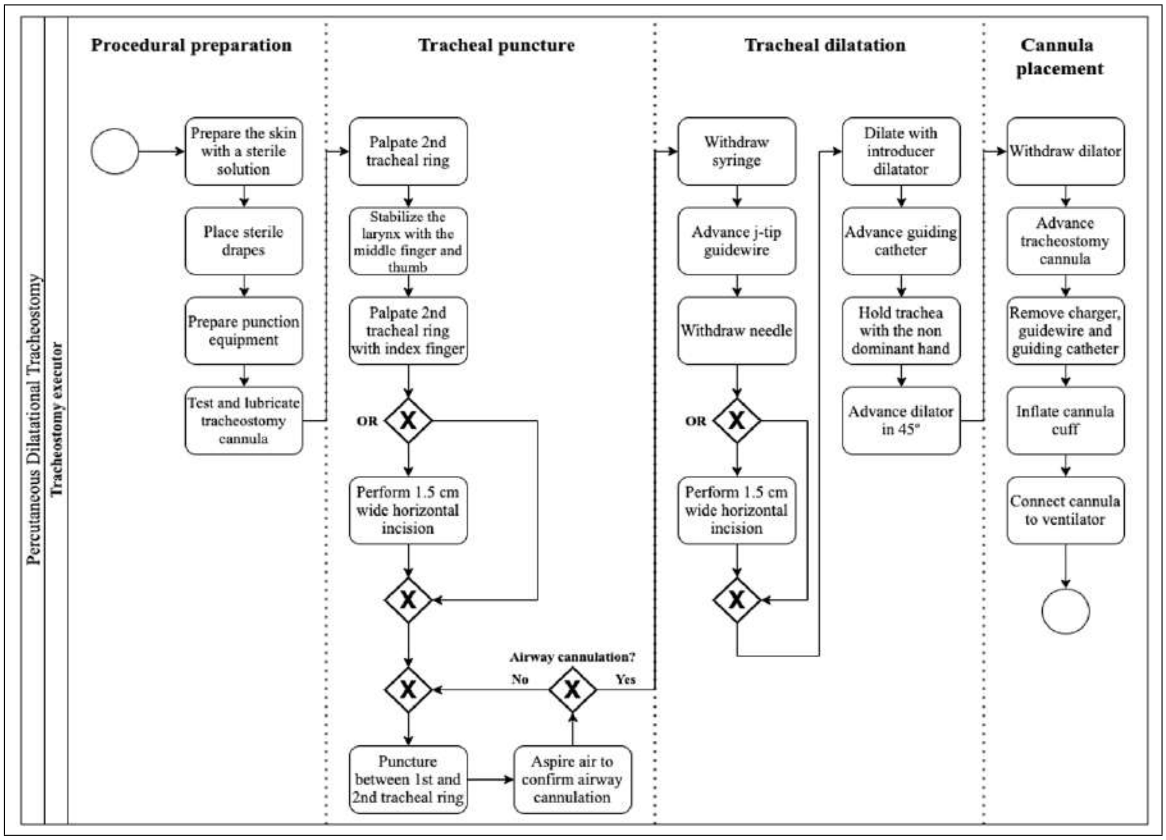


Figure 6.1. Reference PDT process model, based on (de la Fuente et al., 2021).

6.2.2. Data collection

Videos of residents performing PDT training in a low-cost BG-PDT simulation and in a cadaveric model (Kattan et al., 2019) were collected. 8 senior residents (postgraduate year 3) from anesthesiology, emergency medicine, internal medicine or first year intensive care from the Faculty of Medicine of the Pontificia Universidad Católica de Chile participated in the training. Before the evaluation, residents were shown relevant BG-PDT literature and a step-by-step video of the complete procedure performed in the simulation. A total of 56 sessions were performed, 7 sessions per resident, 6 in the simulation center and 1 cadaveric session (session 7).

The videos recorded the execution of each of the steps of the procedure performed by the residents, from the beginning of the procedure, including the preparation of instruments to the connection of the cannula. To generate the event logs, the videos were labelled using the software

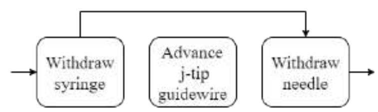
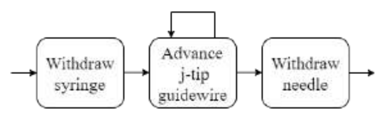
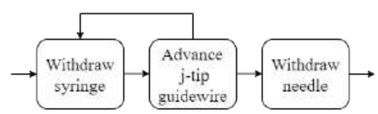
POMElog (Leiva et al., 2019). For each activity executed by the resident, a unique identifier of the execution (Case Id) is stored, in this case an identifier that combines the corresponding resident and a correlative of each training session of the procedure; the executed step (activity); the resident that executed the step (executor); and the start and end of each activity (start and complete timestamps).

6.2.3. Procedure Analysis

6.2.3.1. Process-oriented metrics definition

This paper proposes the use of three process-oriented metrics that allow a top-down analysis. These metrics measure the degree of non-compliance of the execution with the procedural model, showing the differences with the expected execution. Given the sequence of activities $S =< a_1, a_2, ..., a_{n-1}, a_n >$ described by the defined model, and a_k the activity expected at position k in the sequence, Table 6.1 presents the process-oriented metrics to be used in the study. It should be noted that it is assumed that there are no parallel paths in the execution of the procedure; the metrics could be generalized to consider the existence of parallelism. It should be noted that to calculate the deviation metric for each activity, we counted every time the activity was the origin, the destination or it was in between the origin and destination activities of the deviation, according to the procedural model.

Table 6.1. Process-oriented metrics that quantify errors made by residents during training, compared to the model defined for this procedure (PDT).

Name	**Objective**	**Formal Definition**	**Example**
Omission	To record whether a mandatory activity is not performed during the execution.	Count how many times a mandatory activity a_k, $1 \leq k \leq n$, is not performed in the execution.	
Repetition	To record whether an activity is repeated consecutively during execution.	Count how many times an activity a_k, $1 \leq k \leq n$, is followed by the same activity a_k in the execution.	
Deviation	To record whether an activity is performed in the incorrect order.	Count how many times an activity a_k, $2 \leq k \leq n$, is followed by an activity a_j, $1 \leq j < k$, in the execution sequence.	
Aggregated metric	To record the result of all process-oriented metrics for an activity.	The sum of all process-oriented metrics involving an activity a_k	

6.2.3.2. Process mining top-down analysis

Based on the defined model (Figure 6.1) and the data collected, a top-down analysis was developed, starting from a broad perspective, analysing the process as a whole, and then moving to a more detailed level, analysing at the stage level, and then at the activity level. Subsequently, the correlation between the classic metrics of the procedure and the process-oriented metrics was studied.

The top-down methodology used begins with an analysis of the outcome of each resident's process-oriented metric aggregated into a total outcome per session. The Wilcoxon test was used to measure whether the progress between sessions was statistically significant at the group level ($p < 0.05$). Then, the above result is broken down for each of the stages of the procedure. From the previous result, the two stages with the worst results for the process-oriented metrics were selected, and their activities were analyzed according to these metrics. Finally, the evolution of two particular activities during training was analyzed.

6.2.3.3. Classic metrics and process-oriented metrics

The linear correlation between process-oriented metrics and classic metrics was measured to study their behavior as the sessions progressed. To represent the performance of the residents in the classic metrics, the results of the execution time and OSATS metrics from the simulation study (Kattan et al., 2020) were used. These two classic metrics were analyzed in this research since they are not restricted to any specific procedure (Niitsu et al., 2013). Execution time was measured from the moment the resident started the first activity in the model until the last activity was completed. To measure OSATS, the videos were reviewed by two blinded experts. The time metric is considered to improve as it decreases, while OSATS ranges from 1 to 25, with 25 being the maximum expected score. To obtain the correlation values, Pearson's correlation index was used, comparing each process-oriented metric with each classic metric for each procedure execution separately. The result was considered significant with a p-value lower than 0.05.

6.3. Results

6.3.1. Process-oriented metrics at high level

The results show a tendency for the total errors of the residents to decrease as they progress through their training sessions (Figure 6.2), with the exception of the cadaveric session (session 7). In this last session, in particular, the deviations increase with respect to the sixth session with simulator. At the end of the training, the residents performed the procedure without repetitions, and omissions decreased. On the other hand, deviations did not improve significantly over the sessions. For the omission metric, there was a statistically significant decrease in sessions 5, 6 and 7 with respect to session 1 (Wilcoxon, $p < 0.05$). On the other hand, no evidence was found to establish a statistically significant difference in the other metrics.

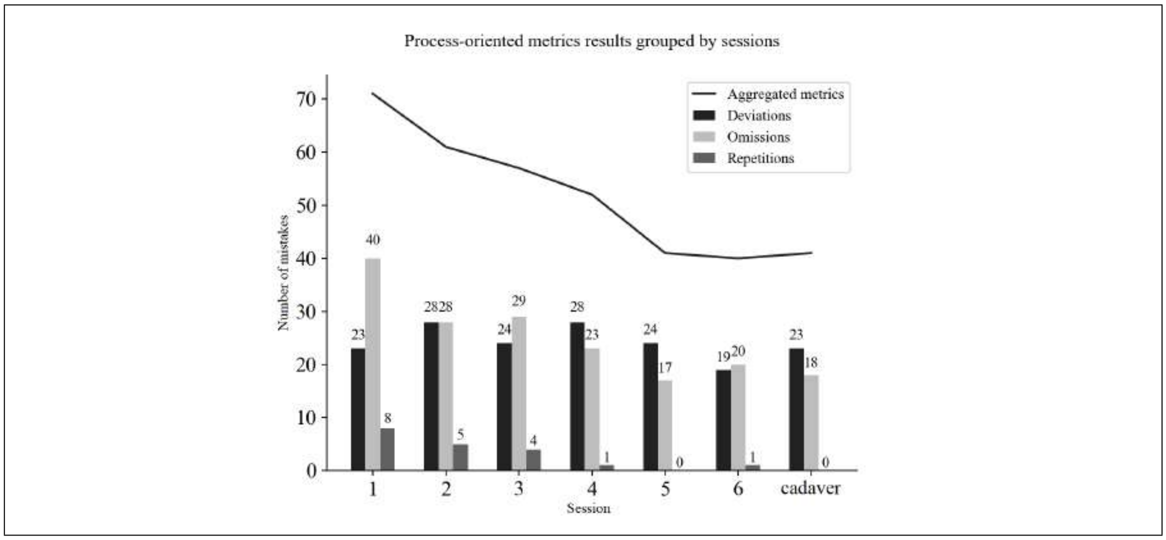


Figure 6.2. Process-oriented metrics results of all students, grouped by session.

6.3.2. Process-oriented metrics by stage

Continuing with the top-down analysis, Figure 6.3 shows how the results of the process-oriented metrics are distributed in each of the 4 stages of the procedure.

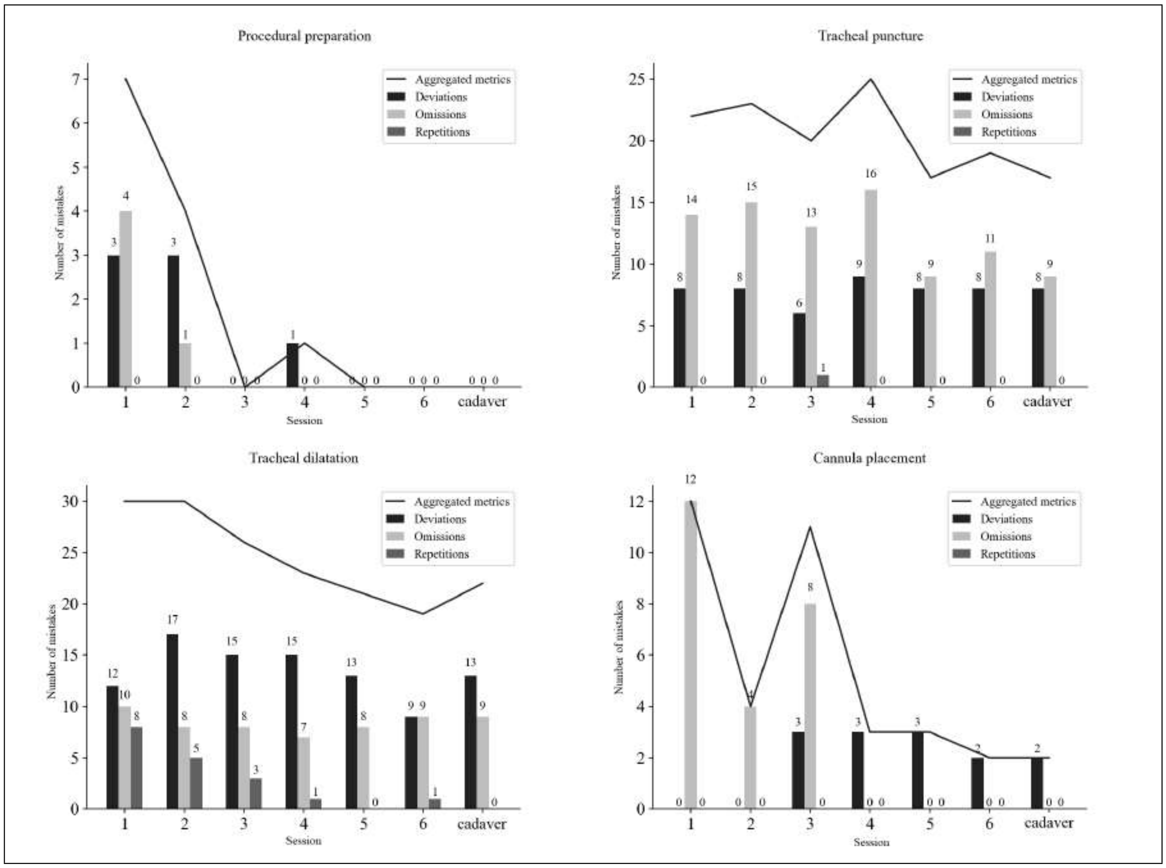


Figure 6.3. Process-oriented metrics results grouped of all students for each stage, grouped by session.

Procedural preparation stage

For the Procedural preparation stage, a maximum of 7 total errors were observed, which occurred in the first session. Throughout the 7 sessions, there were 0 repetitions. After the third session, there were no more omissions, and after the fourth session, there were no more deviations.

Tracheal puncture stage

In the Tracheal puncture stage, the maximum number of errors occurred in the fourth session, with a total of 25 errors. Throughout the 7 sessions, there was only 1 repetition in session 3. Omissions decreased in the last three sessions, but not substantially. Deviations, on the other hand, remained more or less constant throughout the sessions. In all sessions, omissions predominated over the other metrics.

Tracheal dilatation stage

For the Tracheal dilatation stage, the maximum number of errors occurred in the first and second session, reaching a total of 30; then they decreased throughout the sessions, having a slight upturn in the session with cadaver. It is observed that the repetitions decreased as the sessions progressed. However, there was a repetition in session 6, after having observed that they had ceased to occur in session 5. Omissions did not tend to decrease. Deviations and omissions were detected in all sessions. Also, in all sessions the value of the deviation metric predominated over the others.

Cannula placement stage

For the cannula placement stage, the maximum number of errors occurred in session 1, where all the errors corresponded to omissions of procedural activities. In session 3 the omissions ceased, and the deviations started to happen, which were maintained until the last session.

6.3.3. Process-oriented metrics in activities

We proceeded to analyze the activities of the two stages that did not show a clear tendency to improve with the development of the sessions (Tracheal puncture and Tracheal dilatation). Figure 6.5 shows the sum of the results obtained in the process-oriented metrics considering all the sessions and all the residents (56 executions in total) for the selected stages, broken down by activities. This shows which activities can be the most difficult for the residents to learn.

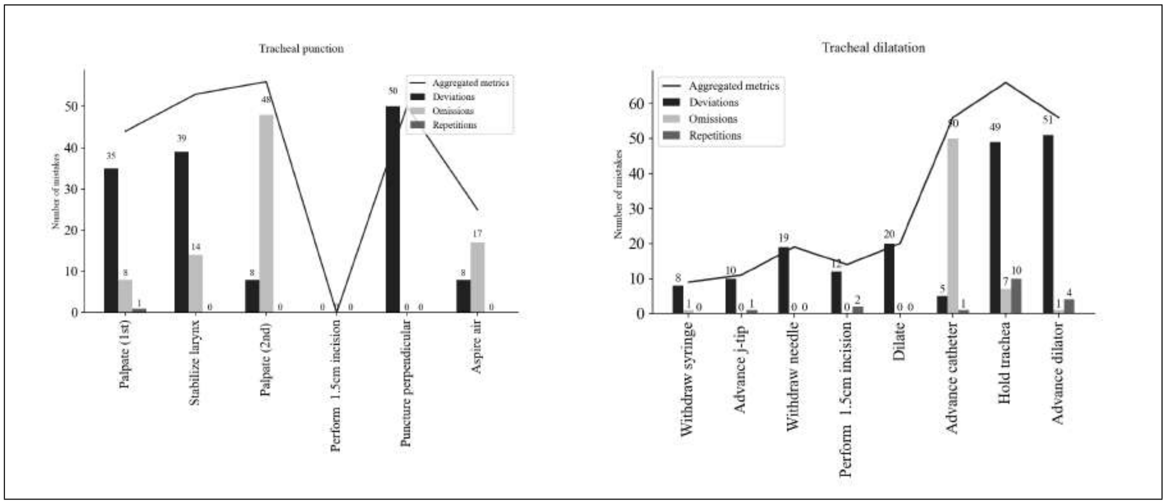


Figure 6.5. Process-oriented metrics results of all students and all sessions for tracheal puncture and tracheal dilatation stages, grouped by activity.

All activities in these model stages have scores greater than 0 on all three process-oriented metrics, except for *Perform 1.5 cm wide horizontal incision*, which is not a mandatory activity and, therefore, has no omissions. Of the 13 activities with errors, 10 are dominated by deviations. In the remaining 3, omissions predominate.

Tracheal puncture

The activities with the highest number of errors in the Tracheal puncture stage are: *Palpate 2nd tracheal ring* (56 aggregated errors; 48 omissions), *Stabilize the larynx with the middle finger and thumb* (53 aggregated errors; 39 deviations) and *Puncture between 1st and 2nd tracheal ring* (50 aggregated errors; 50 deviations).

Tracheal dilatation

The activities with the highest number of aggregated errors in the Tracheal dilatation stage are: *Hold trachea with the non-dominant hand* (66 aggregated errors; 49 deviations), *Advance guiding catheter* (56 aggregated errors; 51 deviations) and *Advance dilator in 45°* (56 aggregated errors; 50 omissions).

6.3.4. Progress of process-oriented metrics in activities

The evolution of two specific activities was analyzed, *Advance dilator in 45° until positioning marks meet*, which belongs to a stage that did not show a relevant improvement during training (Tracheal dilatation) and *Withdraw dilator, leaving guidewire and guiding catheter*, which belongs to a stage that did improve (Cannula placement). Figure 6.7 shows that the first one does not present a clear positive evolution, since it starts with 11 aggregated errors (6 deviations; 1 omission; 4 repetitions) and, as the sessions progress, the deviations tend to be maintained. However, from session 5 onwards, there are no more omissions or repetitions. In contrast, the *Withdraw dilator, leaving guidewire and guiding catheter* activity shows that the residents decreased their errors during training. They start with 4 aggregated errors (2 deviations; 2 omissions) in session 1 and then, from session 5 onwards, they do not present any more errors.

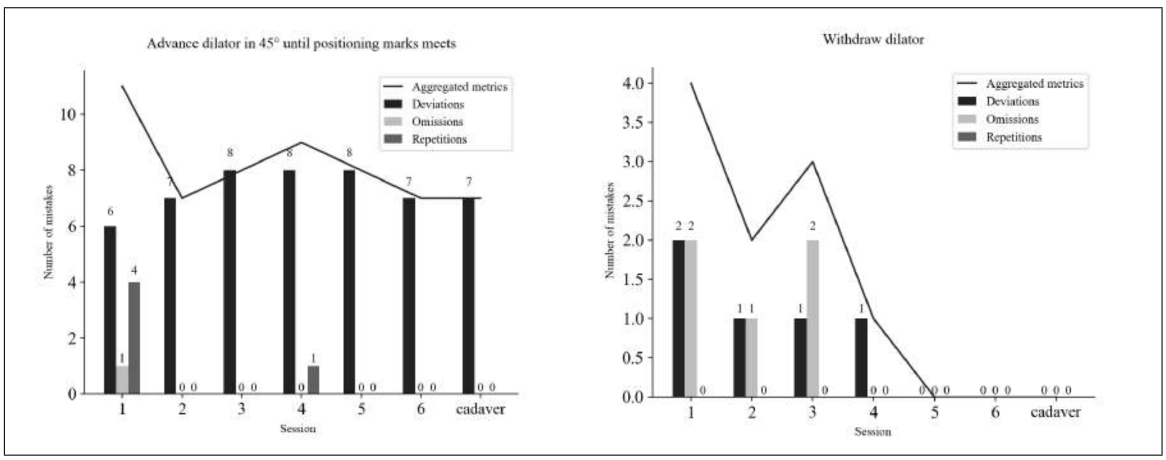


Figure 6.7. Process-oriented metrics results of all students for Advance dilator and Withdraw dilator activities, grouped by session

6.3.5. Comparison between process-oriented metrics and classic metrics

All process-oriented metrics have a negative correlation with the OSATS metric, showing that the reduction of errors in the execution of procedural activities correlates with an increase in the achievement of competence (Table 2). In turn, all process-oriented metrics have a positive correlation with the execution time metric. For all comparison cases, the correlation is statistically significant with a p-value less than the established ($p = 0.05$). Figure 6.9 shows that, like the classic metrics, the process-oriented metrics showed improvement as the sessions progressed.

Table 6.2. Correlation results between classic metrics and process-oriented metrics considering all executions (56). Significance value (p-value) in parentheses.

		Process-oriented metrics		
		Repetition	**Omission**	**Deviation**
Classic metrics	**OSATS**	-0.53 (p < 0.01)	-0.63 (p < 0.01)	-0.45 (p < 0.01)
	Execution time	0.57 (p < 0.01)	0.71 (p < 0.01)	0.54 (p < 0.01)

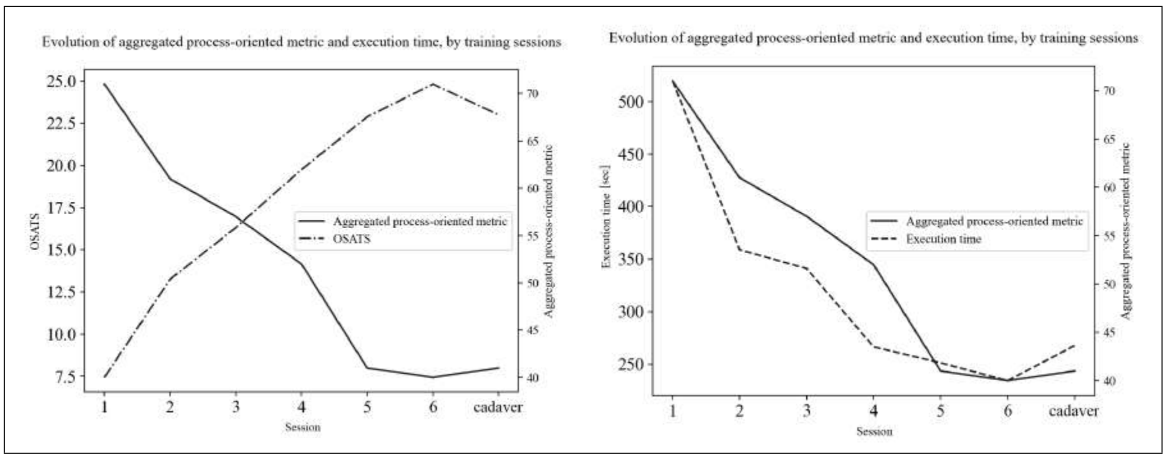


Figure 6.9. Classic metrics versus process-oriented metrics considering all executions (56).

6.4. Discussion

This study proposes the use of process-oriented metrics that allow the execution of the PDT surgical procedure by residents in training sessions to be analyzed using a process perspective. The top-down analysis identified which stages and which activities of the procedure are the most difficult for residents to learn regarding the order of the steps. Overall, process-oriented metrics show similar behavior to classic metrics while providing a higher level of detail to understand how residents evolve during the training process.

Finding the critical points of the procedure is key to design an effective training plan for residents. The cognitive load theory (Sweller, 1988) describes the mechanisms that the working memory uses when processing information. These mechanisms show that, while learning a new procedure, the trainees' working memory can reach its limit and, when overwhelmed, their learning capacity is diminished. This study, through top-down analysis, shows that it is possible to identify the stages and activities in which the residents might be overwhelmed and, consequently, at the end of the course they do not achieve a proficient result according to the process-oriented metrics. At stage level, the stages Tracheal puncture and Tracheal dilatation were identified as those where most procedural errors occur. Then, at the activity level, those activities with the highest number of errors in the stages analyzed in detail are: *Palpate 2nd tracheal ring* (Tracheal puncture), *Advance guiding catheter* (Tracheal dilatation), *Advance dilator in 45°* (Tracheal dilatation) and *Hold trachea with the non-dominant hand* (Tracheal dilatation). Finally, the process-oriented metrics allow the evolution of the activities through the sessions to be reflected, showing which activities do not show progress as the training progresses. For the training process to be effective, all the procedure stages must be well balanced cognitively (McGraw et al., 2019), so as not to overload the residents' working memory. The results show that not all activities and stages have the same cognitive load, which could be generating *"cognitive bottlenecks"* (J. Q. Young et al., 2014), unbalancing the learning process. A new decomposition of activities and stages could have positive effects on student outcome (Nicholls et al., 2016). By limiting the amount of information delivered per stage of the procedure and with the cognitive load better distributed, a more balanced model of the procedure could be conveyed (J. Q. Young et al., 2014) improving the learning experience of the residents. For instance, it would be possible to review the current decomposition and create smaller activity groups than the stages, having in mind the difficulty and sense of the activities involved.

Feedback in the context of medical education is defined as the delivery of specific information on the performance of a trainee compared to an expert level, given with the intent to improve the trainee's performance (van de Ridder et al., 2008). For feedback to be optimal, it should be as specific as possible, since, when an instructor points out specific errors in performance, trainees are less likely to perceive the information as personal criticism (Ghaderi & Farrell, 2020), allowing the trainee to focus on specific points to improve in the performance of the procedure. The results of process-oriented metrics deliver standardized information that, from the highest to the lowest level, would allow residents to understand in which stages/activities of the specific procedure they are failing and how they are progressing in those stages/activities during training. In addition, the feedback provided by process-oriented metrics would be less emotionally charged and stressful than a global evaluation itself, given the impartiality with which it is presented. Minimizing the burden of affective factors that are directly related to cognitive load could optimize procedural learning (Szulewski et al., 2021). In addition to the above, process-oriented feedback has been well received by residents (Lira et al., 2019), who state that the standardization of the process together with concrete feedback improves their training experience.

To the best of our knowledge, while there are studies on tracheostomy training (Kristensen et al., 2015; Nakai et al., 2018), they are oriented toward describing the technique of the procedure rather than discussing the most effective way to teach it or provide feedback to residents. This study proposes the use of process-oriented metrics to reinforce the analysis of this procedure and thus complement traditional evaluation methods, increasing the information generated for both instructors and residents. These metrics do not contradict the classic metrics, since, at a global level, these metrics presented an evolution similar to the classic metrics used in the simulation study (Kattan et al., 2020). Throughout the training sessions, the process-oriented metrics improved as well as the classic metrics. Even for session 7, in the switch to cadaver, both get worse. Coherently, a statistically significant correlation was observed between each of the process-oriented metrics with the classic metrics. However, process-oriented metrics are capable of identifying that there are activities and stages of the process in which there is still room for improvement, according to the process perspective, which is not detected by classic metrics. For session 6, the classic metrics are already proficient in OSATS and execution time, while among the process-oriented metrics, deviations and omissions still have non-optimal values. This makes us observe, as is pointed out by de la Fuente et al. (2020), that the process perspective helps to find hidden information that is not reviewed by traditional evaluation methods or by the observation of the instructors themselves.

There are certain limitations to this study. In the first place, the model used does not represent the complications that a real execution could have. In addition to the above, no analysis has been carried out on the execution of experts, so the conception of proficiency could vary if, when observing the execution of an expert under real situations, it is concluded that it is good in classical metrics, but not in process-oriented metrics. On the other hand, in the definition of the model used in the study, several of the activities of the original model were left out, excluding the bronchoscopist from the procedure. This was due to the fact that this study focuses on the learning of procedural skills, while the bronchoscopist's activities are monitoring activities, therefore, considered secondary for this study.

In conclusion, the process-oriented metrics capture new information to analyze the performance of the residents in the execution of the PDT procedure, which allows us to provide them with adequate and more detailed feedback as they progress through the training sessions, both at the global level and at the stage and activity levels. These process- oriented metrics showed a statistically significant correlation with the classic metrics, so it is promising to propose them as a tool to complement the analysis of the results provided by the evaluation methods currently used. Additionally, through these process-oriented metrics, instructors can be made aware of the weak points of procedural learning. With this information, they can evaluate whether instructional resources should be redeployed to make the teaching process more efficient.

Chapter 7

Conclusions and Future Work

7. CONCLUSIONS AND FUTURE WORK

At first sight, it seems that there was no relationship between process mining and medical education. There were neither data nor clearly established artifacts showing the possibility of linking both topics and producing improvements in procedural skills training. This thesis shows the opposite, since we were able to shape the POME approach thanks to efforts coming from medicine, education and engineering. Such efforts allowed the development of POME artifacts to improve the education of future physicians. As mentioned in Herzlinger (2006), innovations in healthcare are in three areas: the way healthcare is consumed, the use of technology to improve patients health and new business models. In this thesis, we developed POME artifacts through technology to improve medical education of procedural skills, and finally, the patients health.

Shaping the POME approach is at the heart of supporting instructors through the procedural skill teaching cycle. Currently, it seems that existing instruments consider a relevant aspect of procedural skills superficially, which is the control-flow perspective (i.e., the sequence of steps). As this thesis proposes, the POME approach allows instructors to teach a procedural skill as if it were a process by means of POME artifacts, which are solutions that consider the control-flow perspective and can be used in specific stages of the procedural skill teaching cycle.

7.1. Contributions

This thesis describes five contributions to frame the POME approach. First, a systematic literature review to identify existing instruments and strategies considering the control-flow perspective in the medical education literature was performed. It makes explicit the necessity of developing POME artifacts to include the control-flow perspective across the procedural skill teaching cycle. Second, a method to develop POME artifacts to support instructors in each stage of the procedural skill teaching cycle was described, which establishes the stages to create POME artifacts. Third, a POME instrument to let instructors know the learning deficiencies of residents after a training session was designed and validated to support instructors in the deficiencies identification stage. Fourth, a POME learning curve was built to understand the residents performance across all the sessions of a training course. Fifth, POME metrics to obtain a detailed performance analysis of residents thoughout a training course were designed and calculated. Both POME learning curve and metrics were built to support instructors in the performance stage of the procedural skill teaching cycle, providing them with information about control-flow performance.

In the paragraphs below descriptions of each contribution are presented:

7.1.1. Systematic Literature Review

After systematically searching the literature for articles that explicitly show the teaching and assessment of the sequence of steps of surgical procedures, chapter 2 shows that this has not been a well-studied topic. Only 9 articles were found, in which the use of videos as a strategy to teach the sequence, instruments similar to non-standardized scales to assess the sequence of steps, and outcomes such as adherence to a predefined sequence and omission of steps to evaluate the learning of residents were found. We also proposed that this aspect of surgical procedures (i.e., the sequence of steps) could be highlighted and incorporated into procedural skills training through the use of process models depicting a surgical procedure.

7.1.2. POME method

Chapter 3 details the stages of the method to generate POME artifacts that support medical education based on process mining. In this thesis, the method was applied to develop POME artifacts to support instructors in their tasks of detecting learning deficiencies and understanding residents' performance throughout the training course. The method consists of 3 stages: "Model stage", in which a consensus is generated among experts on the sequence of steps that the process should be executed; "Data stage", in which the residents are recorded performing the surgical procedure and then the videos are labeled through POMElog (Leiva et al., 2019) to obtain data on the sequence of steps performed; and finally "Analysis stage", stage in which the POME artifact is developed for the task needed in medical education, based on process mining algorithms.

7.1.3. Control-flow information for instructors

In this thesis we present POME artifacts developed under the POME method: a POME instrument for instructors to know residents' deficiencies in control-flow learning, a learning curve comparing residents' performance with an ideal execution of the procedure, and POME metrics to determine residents' control-flow performance.

Chapter 4 presents the POME instrument for instructors. The first contribution is a taxonomy to classify the steps of a surgical procedure according to their nature (prepare materials and patient, identify a structure, perform an action, verify that the action has been performed correctly). The second contribution is the instrument itself, which was developed based on the taxonomy. Both artifacts were validated with a group of instructors with experience in teaching the central venous catheter installation, using a mixed methodology. The procedure mentioned was used as a case study for the development of both artifacts. After validation, it was concluded that the instrument is understandable for the instructors and that it can be useful in their daily teaching tasks.

Chapter 5 shows the construction of a POME learning curve of a surgical procedure. In particular, data from a bronchoscopy-guided percutaneous tracheostomy course were used. Using a similarity metric, it was possible to compare residents' performances with the ideal performance according to the process model that represents it. We found that there is indeed a learning curve of the control-flow of the surgical procedure. We observed that it reaches a plateau, that despite advancing in similarity towards the end of the course the residents did not perform the procedure identically to the sequence described by the model, and we also observed that when disaggregated by stages of the procedure, some stages were learned better than others. With this approach, instructors of a percutaneous tracheostomy course can learn which stages and steps are more complex, and thus adapt their courses to improve resident learning.

Finally, Chapter 6 presents a top-down analysis to further detail the results found in Chapter 5. POME metrics were developed to quantify control-flow errors made by residents in a percutaneous tracheostomy course. These metrics consist of omission, repetition and deviation from the order defined by the model. With this analysis it was possible to analyze which were the stages and steps in which the residents made most mistakes, as well as to validate the metrics through correlation analysis with classic medical education metrics (execution time and OSATS), finding statistically significant correlations.

7.2. Concluding statements

After all the contributions, three statements can be concluded. The first one is that shaping the POME approach fosters the explicit inclusion of the control-flow perspective across the procedural skill teaching cycle. This need was derived from the systematic literature review results, and it is partially addressed by the POME artifacts developed in this thesis (artifacts for two of five stages were developed: deficiencies identification and performance stages). After developing the POME instrument, the second conclusion is the need of process knowledge but also medical sense to build novel POME artifacts that make sense to instructors. A third conclusion is that, based on the POME learning curve and the POME metrics, the residents do not perform the procedural skill as they are supposed to (i.e. according to a consensus process model depicting the surgical procedure), because even at the end of training courses they still commit control-flow mistakes.

Therefore, considering the technologies available in the submission year of this thesis and the POME artifacts presented before, the POME approach in action (i.e, using it to teach a procedural skill) would have the following steps:

1. As suggested in the POME method, a process model describing the sequence of steps and decision points to perform the surgical procedure involved should be generated through consensus.
2. Classify the steps of the consensus process model using the taxonomy, i.e., categorize each step as preparation, identification, action, control or other steps.
3. Prior to the start of the training course, make sure of having all the implements needed to perform the procedure, the ethical approvals to record the residents while performing the procedure and the setting ready to record the procedures.
4. Introduce the surgical procedure to the class as it is commonly done. Dedicate time in the introductory session to show the consensus process model to the residents, focusing on the sequence of steps, its classification and the decision points.
5. Proceed to run the first session. Once all the residents completed the session performing the procedure, the videos can be tagged by an expert in the procedure using POMElog. Build the POME instrument using that data, and with the information presented there

at the course level decide whether a teaching strategy change is needed. Use the information provided for each student to give personalised feedback to each student prior to the next session, pointing out the mistakes and correct sequence using the consensus process model.

6. Repeat the step 5 for all the remaining sessions.
7. Once the course is finished, build the POME learning curve to have a sense of the overall course performance, and calculate the POME metrics to obtain detailed information about control-flow mistakes committed throughout the course. Consider this information to assess the suitability of the current teaching strategy for future training courses.

The POME approach in action, as presented, it is certainly subject to improvements to enable its use in a common daily basis. Hence, limitations that need to be tackled to enhance the approach are presented.

7.3. Limitations

The limitations of the research conducted, separating them in limitations of the thesis itself and when implementing the POME approach in procedural skills training, are described below.

7.3.1. Research presented in this thesis

Regarding the Systematic Literature Review, one limitation is that we found only nine articles that passed the eligibility criteria we defined. This number of articles can be improved reviewing the string we used to retrieve articles, with the aim of establishing strong conclusions about the trends in teaching and assessing the sequence of steps in procedural skills training.

At the moment, a relevant limitation of the POME method is the time spent in data collection, which is relevant to put in action the POME approach. Reviewing the videos to generate the data is time-consuming, since a large amount of videos needs to be reviewed and this task has to be done by an expert in the procedure. This is an important point, as it is known that medical instructors are facing time constraints for teaching tasks across the world (Roshetsky et al., 2013).

As stated in Chapter 4, the POME instrument designed was validated with three experts in the procedure involved. This is certainly a low number, but it was enough to uncover possible improvements in its design and in the way to show control-flow information. Also, the POME instrument is useful when it is implemented to analyze simple procedures that contains a few control-flow gateways, i.e., improvements are needed to provide the information generated in an easy-to-understand way, especially when the procedure is long and complex in terms of control-flow.

When it comes to the POME learning curve, the similarity metric we used to build it showed to be effective to describe the control-flow progress through the sessions. However, a study on choosing the best function to calculate the similarity metric would help to make the learning curve preciser. We used the model trace that maximized the similarity with the resident trace (the model contained four possible traces). It is possible that the similarity average or the median of all the traces acceptable by the process model, instead of the maximum similarity, produces a more realistic learning curve.

Finally, regarding the POME metrics, they are useful to describe the progress of novices in a simulation context. It needs to be tested whether the metrics are useful to describe an expert performance, and also whether they are useful to describe novices performance in a real setting, due to its inherent complexity.

7.3.2. Implementing the POME approach

It is important to mention that the POME approach focus on the control-flow perspective of procedural skills, and it is not intended to replace the current strategies in place in medical education. The POME artifacts presented are useful to provide objective information about mistakes and performance regarding the control-flow aspect, but they provide neither precise nor accurate information about other aspects of procedural skills such as appropriate assistance usage and instrument handling (McKinley et al., 2008).

Another issue to implement the POME approach is that generating the POME instrument, building the POME learning curve and calculating the POME metrics is not automatic by now. This means that there is no such system to give the data as input and automatically obtain the

POME instrument, learning curve and metrics. Building such a system would enable the use of the POME approach in a daily basis, and support instructors with information in a timely fashion. Note that such system needs to be tested with instructors, to ensure the adoption of the POME artifacts in the daily medical education practice.

Lastly, this thesis proposes POME artifacts for two of the five stages of the procedural skill teaching cycle. Therefore, the POME approach can not be fully implemented across the whole teaching cycle, as there is a need for building and validating POME artifacts for the other three stages. Ideas on how to tackle this limitation are presented in the next section.

7.4. Future work

To tackle the limitations presented in the previous section, guidelines on future research are described below.

7.4.1. Related to performance and deficiencies identification stages

Since the main limitation of the POME method is its time-consuming feature for domain experts, it is advisable to make efforts on extracting event logs from videos automatically (Kratsch, König, & Röglinger, 2022). For such end, recording new videos and establishing the best setting to collect them is a need to be addressed. In this way, it would be easier to train deep learning models to automatically detect the activities.

In relation to the POME instrument, a study on evaluating the understandibility of the instrument with different versions of it (e.g. comparing the one presented in this thesis and a new one using natural language) would help to validate the usefulness of the POME instrument. Also, it would help to recruit more instructors to make possible statistical inferences. Another interesting research is the implementation of the instrument in a different procedure with more complexity in terms of control-flow. Such study would help to scale this solution, and to produce new ways to present the information to instructors.

Regarding the POME learning curve, a study on how to calculate the similarity metric (i.e. which function to use: the maximum between the possible traces, the average, the median, among others) and its effect on building the learning curve would precise the methodology to calculate it.

Lastly, and similar to the POME instrument, the usefulness of the POME metrics would be demonstrated after its implementation in a more complex procedure, to evaluate its scalability and definition's suitability for such procedures. Besides, a study characterizing experts' performances using these metrics would help to establish quantitatively the expert level of competence, as well as a study of novices performing the procedure in a real setting.

7.4.2. Shaping the other stages

The POME artifacts presented in this thesis are proven to be useful for the deficiencies identification and performance stages of the procedural skill teaching cycle. Hence, research is needed to build and validate POME artifacts for the training, assessment and feedback stages, and therefore implement the POME approach in all the stages.

For the training stage, it is needed to find the most suitable and effective strategy to teach the sequence of steps. This strategy should consider the cognitive load issue that can make the learning by residents difficult. One option is to divide the consensus process model in stages teaching the procedure by stages and from the easiest to the hardest stage, thus avoiding the cognitive overload (J. Q. Young et al., 2014; Sewell et al., 2019). Another option to tackle this issue is showing the next step and decision points on a monitor, based on the consensus process model of the procedure being taught (Lehmann et al., 2016). Also, it is possible to combine the POME artifacts presented in this thesis to develop a complete strategy to teach a course (as the seven stages proposed previously for the POME approach). However, its contribution to improvements in learning still needs to be proven. In other words, a demonstration that a teaching strategy composed by POME artifacts improves the learning of the sequence of steps is a gap to be closed.

For the assessment stage, a POME artifact to determine the learning of a sequence of steps needs to be created. As pointed out in Chapter 2, a few instruments are reported in the literature to assess the control-flow aspect of procedural skills, mainly through the adherence to a standard sequence of steps and counting omissions through checklists and rating scales (see Table 2.3). To avoid human bias and generate a standard instrument to assess learning, a POME artifact based on process mining algorithms can be designed. To this end, it needs to be demonstrated whether the consensus process model reflects an expert performance to use it as reference, and decide how students are going to be classified, e.g., defining a pass/fail score, a threshold of acceptable mistakes or proficiency bands. Also, this POME artifact will have to be validated using the existing frameworks to do so in the medical education literature, in order to show its ability to assess the sequence of steps learning (Borgersen et al., 2018).

For the feedback stage, similar to the training stage, there is a need to find the most suitable and effective strategy to provide feedback about the sequence of steps using POME artifacts. Efforts have been previously done in this regard (Lira et al., 2019), having a good acceptance by the residents. Besides, it is suggested in the fifth step of the POME approach that the POME instrument presented in this thesis can be used to give personalized feedback. It could also be that the POME metrics used for the performance stage are useful to provide feedback after each training session. However, all these approaches to give feedback need to demonstrate its impact in learning. Therefore, designing an intervention comparing a control group and an intervention group being exposed to feedback given using the POME instrument and POME metrics would demonstrate the sequence of steps feedback impact on learning. Also, such intervention should consider the instance of debriefing, in which residents appraise and learn the content (Maestre & Rudolph, 2015).

REFERENCES

Abbott, F., Ortega, M., Bravo, S., Basoalto, R., & Kattan, E. (2021). Can we improve teaching and learning of percutaneous dilatational tracheostomy's bronchoscopic guidance? *SAGE Open Medicine*, *9*.

Abe, T., Madotto, F., Pham, T., Nagata, I., Uchida, M., Tamiya, N., … the ESICM Trials (2018). Epidemiology and patterns of tracheostomy practice in patients with acute respiratory distress syndrome in ICUs across 50 countries. *Critical Care*, *22*(1), 195.

Ahmidi, N., Ishii, M., Fichtinger, G., Gallia, G. L., & Hager, G. D. (2012). An objective and automated method for assessing surgical skill in endoscopic sinus surgery using eye-tracking and tool-motion data. *International Forum of Allergy & Rhinology*, *2*(6), 507–515.

Aitken, R. J. (2012). Lost opportunity cost of surgical training in the Australian private sector. *ANZ journal of surgery*, *82*(3), 145–150.

Anderson, D. D., Long, S., Thomas, G. W., Putnam, M. D., Bechtold, J. E., & Karam, M. D. (2016). Objective Structured Assessments of Technical Skills (OSATS) Does Not Assess the Quality of the Surgical Result Effectively. *Clinical orthopaedics and related research*, *474*(4), 874–881.

Anderson, L. W., et al. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Apramian, T., Cristancho, S., Watling, C., Ott, M., & Lingard, L. (2015). Thresholds of Principle and Preference: Exploring Procedural Variation in Postgraduate Surgical Education. *Academic Medicine*, *90*(11 Suppl), S70–6.

Apramian, T., Cristancho, S., Watling, C., Ott, M., & Lingard, L. (2016). "They Have to Adapt to Learn": Surgeons' Perspectives on the Role of Procedural Variation in Surgical Education. *Journal of Surgical Education*, *73*(2), 339–347.

Aragon, C. E., & Zibrowski, E. M. (2008). Does Exposure to a Procedural Video Enhance

Preclinical Dental Student Performance in Fixed Prosthodontics? *Journal of Dental Education*, *72*(1), 67–71.

Aydin, A., Fisher, R., Khan, M. S., Dasgupta, P., & Ahmed, K. (2017). Training, assessment and accreditation in surgery. *Postgraduate Medical Journal*, *93*(1102), 441.

Azari, D., Greenberg, C., Pugh, C., & Wiegmann, Douglas Radwin, R. (2019). In Search of Characterizing Surgical Skill. *Journal of Surgical Education*, *76*(5), 1348–1363.

Balayla, J., Bergman, S., Ghitulescu, G., Feldman, L. S., & Fraser, S. A. (2012). Knowing the operative game plan: a novel tool for the assessment of surgical procedural knowledge. *Canadian Journal of Anesthesia/Journal canadien d'anesthésie*, *55*(4), S158–S162.

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction*, *24*(6), 574–594.

Barsuk, J. H., McGaghie, W. C., Cohen, E. R., O'Leary, K. J., & Wayne, D. B. (2009). Simulation-based mastery learning reduces complications during central venous catheter insertion in a medical intensive care unit. *Critical Care Medicine*, *37*(10), 2697–2701.

Berman, A. C. (2015). Good teaching is good teaching: A narrative review for effective medical educators. *Anatomical Sciences Education*, *8*(4), 386–394.

Bloom, B. S., & Krathwohl, D. R. (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals by a committee of college and university examiners. In *Handbook i: Cognitive domain.* New York: Longman.

Bogarín, A., Cerezo, R., & Romero, C. (2018). A survey on educational process mining. *WIREs Data Mining and Knowledge Discovery*, *8*(1), e1230.

Borgersen, N. J., Naur, T. M. H., Sørensen, S. M. D., Bjerrum, F., Konge, L., Subhi, Y., & Thomsen, A. S. S. (2018). Gathering Validity Evidence for Surgical Simulation. *Annals of Surgery*, *267*(6), 1063–1068.

Bramer, W. M., Rethlefsen, M. L., Kleijnen, J., & Franco, O. H. (2017). Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Systematic Reviews*, *6*(1), 245.

Brenner, M., Hoehn, M., Pasley, J., Dubose, J., Stein, D., & Scalea, T. (2014). Basic endovascular skills for trauma course. *Journal of Trauma and Acute Care Surgery*, *77*(2), 286–291.

Brooke, J. (1996). SUS: a 'quick and dirty' usability scale. In *Usability evaluation in industry* (Vol. 189).

Carmona, J., van Dongen, B., Solti, A., & Weidlich, M. (2018). *Conformance Checking*. Springer International Publishing.

Chapman, D. M., Marx, J. A., Honigman, B., Rosen, P., & Cavanaugh, S. H. (1994). Emergency Thoracotomy: Comparison of Medical Student, Resident, and Faculty Performances on Written, Computer, and Animal-model Assessments. *Academic Emergency Medicine*, *1*(4), 373–381.

Cheung, J. J. H., Kulasegaram, K. M., Woods, N. N., Moulton, C.-a., Ringsted, C. V., & Brydges, R. (2018). Knowing How and Knowing Why: testing the effect of instruction designed for cognitive integration on procedural skills transfer. *Advances in Health Sciences Education*, *23*(1), 61–74.

Clark, R. E., Pugh, C. M., Yates, K. A., Inaba, K., Green, D. J., & Sullivan, M. E. (2012). The Use of Cognitive Task Analysis to Improve Instructional Descriptions of Procedures. *Journal of Surgical Research*, *173*(1), e37-e42.

Cohen, W. W., Ravikumar, P., & Fienberg, S. E. (2003). A Comparison of String Distance Metrics for Name-Matching Tasks. *Proceedings of the 2003 International Conference on Information Integration on the Web*, 73–78.

Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., . . . Hatala, R. (2013). Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis. *Medical Teacher*, *35*(1), e867–e898.

Cook, D. A., & Reed, D. A. (2015). Appraising the quality of medical education research methods: the Medical Education Research Study Quality Instrument and the

Newcastle-Ottawa Scale-Education. *Academic Medicine*, *90*(8), 1067–1076.

Corvetto, M. A., Pedemonte, J. C., Varas, D., Fuentes, C., & Altermatt, F. R. (2017). Simulation-based training program with deliberate practice for ultrasound-guided jugular central venous catheter placement. *Acta Anaesthesiologica Scandinavica*, *61*(9), 1184–1191.

Crebbin, W., Beasley, S. W., & Watters, D. A. K. (2013). Clinical decision making: how surgeons do it. *ANZ Journal of Surgery*, *83*(6), 422–428.

da Costa, G. O. F., Rocha, H. A. L., de Moura Júnior, L. G., & Medeiros, F. d. C. (2018). Taxonomy of educational objectives and learning theories in the training of laparoscopic surgical techniques in a simulation environment. *Revista do Colegio Brasileiro de Cirurgioes*, *45*(5), e1954.

de la Fuente, R. (2018). *Entrenamiento en Destrezas Procedurales: Oportunidades desde el análisis de Minería de Procesos* (Unpublished doctoral dissertation). Pontificia Universidad Católica de Chile.

de la Fuente, R., Fuentes, R., Munoz-Gama, J., Dagnino, J., & Sepúlveda, M. (2020). Delphi Method to Achieve Clinical Consensus for a BPMN Representation of the Central Venous Access Placement for Training Purposes. *International Journal of Environmental Research and Public Health*, *17*(11).

de la Fuente, R., Fuentes, R., Munoz-Gama, J., Riquelme, A., Altermatt, F. R., Pedemonte, J., … Sepúlveda, M. (2020). Control-flow analysis of procedural skills competencies in medical training through process mining. *Postgraduate Medical Journal*, *96*(1135), 250–256.

de la Fuente, R., Kattan, E., Munoz-Gama, J., Puente, I., Navarrete, M., Kychenthal, C., … Sepúlveda, M. (2021). Development of a comprehensive Percutaneous Dilatational Tracheostomy process model for procedural training: A Delphi-based experts consensus. *Acta Anaesthesiologica Scandinavica*, *65*(2), 244–256.

Dempsey, G. A., Grant, C. A., & Jones, T. M. (2010). Percutaneous tracheostomy: a 6 yr prospective evaluation of the single tapered dilator technique. *British Journal of Anaesthesia*, *105*(6), 782–788.

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*(4), 401–409.

Dias, R. D., Ngo-Howard, M. C., Boskovski, M. T., Zenati, M. A., & Yule, S. J. (2018). Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. *The British Journal of Surgery*, *105*(5), 491–501.

Díaz-Regañón, G., Miñambres, E., Ruiz, A., González-Herrera, S., Holanda-Peña, M., & López-Espadas, F. (2008). Safety and complications of percutaneous tracheostomy in a cohort of 800 mixed ICU patients. *Anaesthesia*, *63*(11), 1198–1203.

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). *Fundamentals of Business Process Management* (Second ed.). Springer, Berlin, Heidelberg. doi: 10.1007/978-3-662-56509-4

Epstein, R. M. (2007). Assessment in Medical Education. *New England Journal of Medicine*, *356*(4), 387–396.

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. *JAMA*, *287*(2), 226–235.

Fecso, A. B., Szasz, P., Kerezov, G., & Grantcharov, T. P. (2017). The Effect of Technical Performance on Patient Outcomes in Surgery. *Annals of Surgery*, *265*(3), 492–501.

Fernández-Llatas, C., Meneu, T., Traver, V., & Benedi, J.-M. (2013). Applying Evidence-Based Medicine in Telehealth: An Interactive Pattern Recognition Approximation. *International Journal of Environmental Research and Public Health*, *10*(11).

Frank, J. R., Snell, L. S., Cate, O. T., Holmboe, E. S., Carraccio, C., Swing, S. R., ... Harris, K. A. (2010). Competency-based medical education: theory to practice. *Medical Teacher*, *32*(8), 638–645.

Fraser, K. L., Ayres, P., & Sweller, J. (2015). Cognitive Load Theory for the Design of Medical Simulations. *Simulation in Healthcare*, *10*(5), 295–307.

Ghaderi, I., & Farrell, T. M. (2020). Toward effective feedback: From concept to practice. *Surgery*, *167*(4), 685–688.

Golden, A., Alaska, Y., Levinson, A. T., Davignon, K. R., Lueckel, S. N., Lynch, K. A. J., … Kobayashi, L. (2020). Simulation-Based Examination of Arterial Line Insertion Method Reveals Interdisciplinary Practice Differences. *Simulation in Healthcare*, *15*(2), 89–97.

Graneheim, U. H., & Lundman, B. (2004). Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. *Nurse Education today*, *24*(2), 105–112.

Grantcharov, T. P., & Reznick, R. K. (2008). Teaching procedural skills. *BMJ*, *336*(7653), 1129–1131.

Green, J. L., Suresh, V., Bittar, P., Ledbetter, L., Mithani, S. K., & Allori, A. (2019). The Utilization of Video Technology in Surgical Education: A Systematic Review. *The Journal of Surgical Research*, *235*, 171–180.

Guerlain, S., Green, K. B., LaFollette, M., Mersch, T. C., Mitchell, B. A., Poole, G. R., … Chekan, E. G. (2004). Improving surgical pattern recognition through repetitive viewing of video clips. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *34*(6), 699–707.

Hales, B., Terblanche, M., Fowler, R., & Sibbald, W. (2008). Development of medical checklists for improved quality of patient care. *International Journal for Quality in Health Care*, *20*(1), 22–30.

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, *32*(4), 1008–1015.

Herzlinger, R. E. (2006). Why innovation in health care is so hard. *Harvard Business Review*, *84*(5), 58–66.

Higgs, J., Jones, M., Loftus, S., & Christensen, N. (Eds.). (2008). *Clinical reasoning in the health professions* (3rd ed.). Elsevier.

Hoffman, R. R. (2008). Human Factors Contributions to Knowledge Elicitation. *Human Factors*, *50*(3), 481–488.

Hosogi, H., Obama, K., Tsunoda, S., Hisamori, S., Nishigori, T., Tanaka, E., … Sakai, Y. (2021). Educational application of intraoperative records from an energy device in laparoscopic gastrectomy: a preliminary report. *Surgery Today*, *51*(5), 829–835.

Huang, G. C., McSparron, J. I., Balk, E. M., Richards, J. B., Smith, C. C., Whelan, J. S., ... Smetana, G. W. (2016). Procedural instruction in invasive bedside procedures: a systematic review and meta-analysis of effective teaching approaches. *BMJ Quality & Safety*, *25*(4), 281–294.

Kattan, E., de la Fuente, R., Putz, F., Vera, M., Corvetto, M., Inzunza, O., ... Bravo, S. (2020). Simulation-Based Mastery Learning of Bronchoscopy-Guided Percutaneous Dilatational Tracheostomy: Competency Acquisition and Skills Transfer to a Cadaveric Model. *Simulation in Healthcare*.

Kattan, E., Vera, M., Putz, F., Corvetto, M., De la Fuente, R., & Bravo, S. (2019). Design and Evaluation of a Low-Cost Bronchoscopy-Guided Percutaneous Dilatational Tracheostomy Simulator. *Simulation in Healthcare*, *14*(6), 415–419.

Khan, K., & Ramachandran, S. (2012). Conceptual framework for performance assessment: Competency, competence and performance in the context of assessments in healthcare – deciphering the terminology. *Medical Teacher*, *34*(11), 920-928.

Kirkpatrick, D. (2007). *The Four Levels of Evaluation: Measurement and Evaluation* (A. S. for Training, Ed.). American Society for Training & Development.

Kratsch, W., König, F., & Röglinger, M. (2022). Shedding light on blind spots – Developing a reference architecture to leverage video data for process mining. *Decision Support Systems*, *158*, 113794.

Kristensen, M. S., Teoh, W. H. L., & Baker, P. A. (2015). Percutaneous emergency airway access; prevention, preparation, technique and training. *British Journal of Anaesthesia*, *114*(3), 357–361.

Lalys, F., & Jannin, P. (2014). Surgical process modelling: a review. *International Journal of Computer Assisted Radiology and Surgery*, *9*(3), 495–511.

Lammers, R. L. (2008). Learning and Retention Rates after Training in Posterior Epistaxis Management. *Academic Emergency Medicine*, *15*(11), 1181–1189.

Lammers, R. L., Davenport, M., Korley, F., Griswold-Theodorson, S., Fitch, M. T., Narang, A. T., ... Robey, W. C. r. (2008). Teaching and assessing procedural skills using simulation: metrics and methodology. *Academic Emergency Medicine*, *15*(11), 1079–1087.

Lateef, F. (2018). Clinical Reasoning: The Core of Medical Education and Practice. *International Journal of Internal and Emergency Medicine*, *1*(2).

Légaré, F., Freitas, A., Thompson-Leduc, P., Borduas, F., Luconi, F., Boucher, A., ... Jacques, A. (2015). The Majority of Accredited Continuing Professional Development Activities Do Not Target Clinical Behavior Change. *Academic Medicine*, *90*(2), 197–202.

Lehmann, R., Seitz, A., Bosse, H. M., Lutz, T., & Huwendiek, S. (2016). Student perceptions of a video-based blended learning approach for improving pediatric physical examination skills. *Annals of Anatomy - Anatomischer Anzeiger*, *208*, 179–182.

Lehmann, R., Thiessen, C., Frick, B., Bosse, H. M., Nikendei, C., Hoffmann, G. F., ... Huwendiek, S. (2015). Improving Pediatric Basic Life Support Performance Through Blended Learning With Web-Based Virtual Patients: Randomized Controlled Trial. *Journal of Medical Internet Research*, *17*(7), e162–e162.

Leiva, L., Munoz-Gama, J., Salas-Morales, J., Galvez, V., Lee, W. L. J., de la Fuente, R., ... Sepúlveda, M. (2019). POMElog: Generating Event Logs from Unplugged Processes. *CEUR Workshop Proceedings*, *2420*, 189–193.

Leotsakos, A., Zheng, H., Croteau, R., Loeb, J. M., Sherman, H., Hoffman, C., ... Munier, B. (2014). Standardization in patient safety: the WHO High 5s project. *International Journal for Quality in Health Care*, *26*(2), 109–116.

Lichtenberger, J. P., Tatum, P. S., Gada, S., Wyn, M., Ho, V. B., & Liacouras, P. (2018). Using 3D Printing (Additive Manufacturing) to Produce Low-Cost Simulation Models for Medical Training. *Military medicine*, *183*(suppl_1), 73–77.

Lira, R., Salas-Morales, J., Leiva, L., Fuente, R. D., Fuentes, R., Delfino, A., ... Munoz-Gama, J. (2019). Process-Oriented Feedback through Process Mining for Surgical Procedures in Medical Training: The Ultrasound-Guided Central Venous Catheter Placement Case. *International Journal of Environmental Research and Public Health*, *16*(11).

Ma, I. W. Y., Brindle, M. E., Ronksley, P. E., Lorenzetti, D. L., Sauve, R. S., & Ghali, W. A. (2011). Use of Simulation-Based Education to Improve Outcomes of Central Venous Catheterization: A Systematic Review and Meta-Analysis. *Academic Medicine*, *86*(9).

Ma, I. W. Y., Zalunardo, N., Pachev, G., Beran, T., Brown, M., Hatala, R., & McLaughlin, K. (2012). Comparing the use of global rating scale with checklists for the assessment of central venous catheterization skills using simulation. *Advances in Health Sciences Education*, *17*(4), 457–470.

Ma, R., Vanstrum, E. B., Nguyen, J. H., Chen, A., Chen, J., & Hung, A. J. (2021). A Novel Dissection Gesture Classification to Characterize Robotic Dissection Technique for Renal Hilar Dissection. *Journal of Urology*, *205*(1), 271–275.

Maestre, J. M., & Rudolph, J. W. (2015). Theories and styles of debriefing: the good judgment method as a tool for formative assessment in healthcare. *Revista Española de Cardiología*, *68*(4), 282-285.

Martin, J. A., Regehr, G., Reznick, R., Macrae, H., Murnaghan, J., Hutchison, C., & Brown, M. (2005). Objective structured assessment of technical skill (OSATS) for surgical residents. *British Journal of Surgery*, *84*(2), 273-278.

Massick, D. D., Powell, D. M., Price, P. D., Chang, S. L., Squires, G., Forrest, L. A., & Young, D. C. (2000). Quantification of the learning curve for percutaneous dilatational tracheotomy. *The Laryngoscope*, *110*(2), 222–228.

McGaghie, W. C. (2015). Mastery Learning: It Is Time for Medical Education to Join the 21st Century. *Academic Medicine*, *90*(11), 1438–1441.

McGraw, R., Chaplin, T., Rocca, N., Rang, L., Jaeger, M., Holden, M., … Fichtinger, G. (2019). Cognitive load theory as a framework for simulation-based, ultrasound-guided internal jugular catheterization training: Once is not enough. *Canadian Journal of Emergency Medicine*, *21*(1), 141–148.

McKinley, R. K., Strand, J., Ward, L., Gray, T., Alun-Jones, T., & Miller, H. (2008). Checklists for assessment and certification of clinical procedural skills omit essential competencies: a systematic review. *Medical Education*, *42*(4), 338–349.

Mead, D., & Moseley, L. (2001). The use of the Delphi as a research approach. *Nurse Researcher*, *8*(4), 4–23.

Menix, K. D. (1996). Domains of learning: interdependent components of achievable learning outcomes. *Journal of Continuing Education in Nursing*, *27*(5), 200–208.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, *339*, b2535.

Morgan, P. J., Cleave-Hogg, D., & Guest, C. B. (2001). A comparison of global ratings and checklist scores from an undergraduate assessment using an anesthesia simulator. *Academic Medicine*, *76*(10), 1053–1055.

Müller, R., & Rogge-Solti, A. (2011). BPMN for Healthcare Processes. *CEUR Workshop Proceedings*, *705*, 65–72.

Munoz-Gama, J. (2016). *Conformance Checking and Diagnosis in Process Mining - Comparing Observed and Modeled Processes*. Springer International Publishing.

Munoz-Gama, J., de la Fuente, R., Sepúlveda, M., & Fuentes, R. (2019). *Conformance Checking Challenge 2019.* 4TU. Centre for Research Data, Delft, Netherlands.

Munoz-Gama, J., Galvez, V., de la Fuente, R., Sepúlveda, M., & Fuentes, R. (2021). Interactive Process Mining for Medical Training. In *Interactive process mining in healthcare* (pp. 233–242).

Munoz-Gama, J., Martin, N., Fernandez-Llatas, C., Johnson, O. A., Sepúlveda, M., Helm, E., … Zerbato, F. (2022). Process Mining for Healthcare: Characteristics and Challenges. *Journal of Biomedical Informatics*.

Nakai, M. Y., Menezes, M. B., Kavabata, N. K., Suehara, A. B., Bertelli, A. A. T., Kikuchi, W., & Gonçalves, A. J. (2018). Percutaneous Tracheostomy: Pearls and Pitfalls, and How to Create a "Hand-On" Training Program Course. In T. P. de Farias (Ed.), *Tracheostomy: A surgical guide* (pp. 93–117). Springer, Cham.

Neumuth, T. (2017). Surgical process modeling. *Innovative Surgical Sciences*, *2*(3), 123–137.

Nicholls, D., Sweet, L., Muller, A., & Hyett, J. (2016). Teaching psychomotor skills in the twenty-first century: Revisiting and reviewing instructional approaches through the lens of contemporary literature. *Medical Teacher*, *38*(10), 1056–1063.

Niitsu, H., Hirabayashi, N., Yoshimitsu, M., Mimura, T., Taomoto, J., Sugiyama, Y., … Takiyama, W. (2013). Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surgery Today*, *43*(3), 271–275.

Oermann, M. H. (1990). Psychomotor skill development. *The Journal of Continuing Education in Nursing*, *21*(5), 202–204.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & Management*, *42*(1), 15–29.

Osterweil, L. J., Conboy, H. M., Clarke, L. A., & Avrunin, G. S. (2019). Process-Model-Driven Guidance to Reduce Surgical Procedure Errors: An Expert Opinion. *Seminars in Thoracic and Cardiovascular Surgery*, *31*(3), 453–457.

Petiot, S., Guinot, P.-G., Diouf, M., Zogheib, E., & Dupont, H. (2017). Learning curve for real-time ultrasound-guided percutaneous tracheostomy. *Anaesthesia Critical Care & Pain Medicine*, *36*(5), 279–283.

Pusic, M. V., Boutis, K., Hatala, R., & Cook, D. A. (2015). Learning curves in health professions education. *Academic Medicine*, *90*(8), 1034–1042.

Raimondi, N., Vial, M. R., Calleja, J., Quintero, A., Cortés, A., Celis, E., … Nates, J. L. (2017). Evidence-based guidelines for the use of tracheostomy in critically ill patients. *Journal of Critical Care*, *38*, 304–318.

Rao, A., Tait, I., & Alijani, A. (2015). Systematic review and meta-analysis of the role of mental training in the acquisition of technical skills in surgery. *The American Journal of Surgery*, *210*(3), 545–553.

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association Between Funding and Quality of Published Medical Education Research. *JAMA*, *298*(9), 1002–1009. doi: 10.1001/jamaa.298.9.1002

Rivilla, A. M., Mata, F. S., González, R. A., Entonado, F. B., de Vicente Rodríguez, P. S., Cruz, M. F., … McWilliam, N. (2009). *Didáctica general*. Pearson Educación.

Rojas, E., Munoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). Process mining in healthcare: A literature review. *Journal of Biomedical Informatics*, *61*, 224–236.

Rolón, E., Chavira, G., Orozco, J., & Soto, J. P. (2015). Towards a Framework for Evaluating Usability of Business Process Models with BPMN in Health Sector. *Procedia Manufacturing*, *3*, 5603–5610.

Roshetsky, L. M., Coltri, A., Flores, A., Vekhter, B., Humphrey, H. J., Meltzer, D. O., & Arora, V. M. (2013). No time for teaching? Inpatient attending physicians' workload and teaching before and after the implementation of the 2003 duty hours regulations. *Academic Medicine*, *88*(9), 1293–1298.

Rudas, M., Seppelt, I., Herkes, R., Hislop, R., Rajbhandari, D., & Weisbrodt, L. (2014). Traditional landmark versus ultrasound guided tracheal puncture during percutaneous dilatational tracheostomy in adult intensive care patients: a randomised controlled trial. *Critical Care*, *18*(5), 514.

Sawyer, T., White, M., Zaveri, P., Chang, T., Ades, A., French, H., ... Kessler, D. (2015). Learn, see, practice, prove, do, maintain: an evidence-based pedagogical framework for procedural skill training in medicine. *Academic Medicine*, *90*(8), 1025–1033.

Scalese, R. J., Obeso, V. T., & Issenberg, S. B. (2008). Simulation technology for skills training and competency assessment in medical education. *Journal of General Internal Medicine*, *23 Suppl 1*(Suppl 1), 46–49.

Scheuerlein, H., Rauchfuss, F., Dittmar, Y., Molle, R., Lehmann, T., Pienkos, N., & Settmacher, U. (2012). New methods for clinical pathways-Business Process Modeling Notation (BPMN) and Tangible Business Process Modeling (t.BPM). *Langenbeck's Archives of Surgery*, *397*(5), 755–761.

Schulz, C. M., Endsley, M. R., Kochs, E. F., Gelb, A. W., & Wagner, K. J. (2013). Situation awareness in anesthesia: concept and research. *Anesthesiology*, *118*(3), 729–742.

Schumann, S., Bühligen, U., & Neumuth, T. (2013). Distance measures for surgical process models. *Methods of Information in Medicine*, *52*(5), 422–431.

Schumann, S., Bühligen, U., & Neumuth, T. (2015). Outcome quality assessment by surgical process compliance measures in laparoscopic surgery. *Artificial Intelligence in Medicine*, *63*(2), 85–90.

Schunk, D. H. (2012). *Learning theories. An educational perspective* (6th ed.). Pearson.

Schut, S., van Tartwijk, J., Driessen, E., van der Vleuten, C., & Heeneman, S. (2020). Understanding the influence of teacher–learner relationships on learners' assessment perception. *Advances in Health Sciences Education*, *25*(2), 441–456.

Seam, N., Lee, A. J., Vennero, M., & Emlet, L. (2019). Simulation Training in the ICU. *Chest*, *156*(6), 1223–1233.

Sewell, J. L., Maggio, L. A., Ten Cate, O., van Gog, T., Young, J. Q., & O'Sullivan, P. S. (2019). Cognitive load theory for training health professionals in the workplace: A BEME review of studies among diverse professions: BEME Guide No. 53. *Medical Teacher*, *41*(3), 256–270.

Shackelford, S., & Bowyer, M. (2017). Modern Metrics for Evaluating Surgical Technical Skills. *Current Surgery Reports*, *5*(10), 24.

Shaker, D. (2018). Cognitivism and psychomotor skills in surgical training: from theory to practice. *International Journal of Medical Education*, *9*, 253–254.

Sharma, S., Sharma, V., Sharma, M., Awasthi, B., & Chaudhary, S. (2015). Formative assessment in postgraduate medical education—Perceptions of students and teachers. *International Journal of Applied & Basic Medical Research*, *5*(4), 66–70.

Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of Educational Research*, *56*(4), 411-436.

Singh, J., & Sing, R. F. (2019). Performance, Long-term Management, and Coding for Percutaneous Dilational Tracheostomy. *Chest*, *155*(3), 639–644.

Skjold-Ødegaard, B., & Søreide, K. (2020). Standardization in surgery: friend or foe? *The British Journal of Surgery*, *107*(9), 1094–1096.

Srinivasan, M., Li, S.-T. T., Meyers, F. J., Pratt, D. D., Collins, J. B., Braddock, C., … Hilty, D. M. (2011). "Teaching as a Competency": competencies for medical educators. *Academic Medicine*, *86*(10), 1211–1220.

Stefanidis, D., Scerbo, M. W., Montero, P. N., Acker, C. E., & Smith, W. D. (2012). Simulator Training to Automaticity Leads to Improved Skill Transfer Compared With Traditional Proficiency-Based Training: A Randomized Controlled Trial. *Annals of*

Surgery, *255*(1), 30–37.

Sullivan, M. E., Yates, K. A., Inaba, K., Lam, L., & Clark, R. E. (2014). The Use of Cognitive Task Analysis to Reveal the Instructional Limitations of Experts in the Teaching of Procedural Skills. *Academic Medicine*, *89*(5), 811–816.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.

Szulewski, A., Howes, D., van Merriënboer, J. J. G., & Sweller, J. (2021). From Theory to Practice: The Application of Cognitive Load Theory to the Practice of Medicine. *Academic Medicine*, *96*(1), 24–30.

ten Cate, T. O., Snell, L., & Carraccio, C. (2010). Medical competence: The interplay between individual ability and the health care environment. *Medical Teacher*, *32*(8), 669-675.

Valsamis, E. M., Chouari, T., O'Dowd-Booth, C., Rogers, B., & Ricketts, D. (2018). Learning curves in surgery: variables, analysis and applications. *Postgraduate Medical Journal*, *94*(1115), 525–530.

van der Aalst, W. (2016). *Process mining: Data science in action*. Springer-Verlag Berlin Heidelberg.

van der Aalst, W. (2019). A practitioner's guide to process mining: Limitations of the directly-follows graph. *Procedia Computer Science*, *164*, 321–328.

van der Aalst, W., Adriansyah, A., & van Dongen, B. (2012). Replaying history on process models for conformance checking and performance analysis. *WIREs Data Mining and Knowledge Discovery*, *2*(2), 182–192.

van de Ridder, J. M. M., Stokking, K. M., McGaghie, W. C., & ten Cate, O. T. J. (2008). What is feedback in clinical education? *Medical Education*, *42*(2), 189–197.

Vedula, S. S., & Hager, G. D. (2017). Surgical data science: The new knowledge domain. *Innovative Surgical Sciences*, *2*(3), 109–121.

Vedula, S. S., Ishii, M., & Hager, G. D. (2017). Objective Assessment of Surgical Technical Skill and Competency in the Operating Room. *Annual Review of Biomedical Engineering*, *19*(1), 301–325.

Wallace, L., Raison, N., Ghumman, F., Moran, A., Dasgupta, P., & Ahmed, K. (2017). Cognitive training: How can it be adapted for surgical education? *The Surgeon*, *15*(4), 231–239.

Walter, A. J. (2006). Surgical Education for the Twenty-first Century: Beyond the Apprentice Model. *Obstetrics and Gynecology Clinics of North America*, *33*(2), 233–236.

Wieringa, R. J. (2014). *Design science methodology for information systems and software engineering*. Springer-Verlag Berlin Heidelberg.

Williams, B. C., Litzelman, D. K., Babbott, S. F., Lubitz, R. M., & Hofer, T. P. (2002). Validation of a global measure of faculty's clinical teaching performance. *Academic Medicine*, *77*(2), 177–180.

Wingfield, L. R., Kulendran, M., Chow, A., Nehme, J., & Purkayastha, S. (2015). Cognitive task analysis: Bringing olympic athlete style training to surgical education. *Surgical Innovation*, *22*(4), 406–417.

WMA Statement on Medical Education (Tech. Rep.). (2017). Ferney-Voltaire, France: World Medical Association.

Yates, K., Sullivan, M., & Clark, R. (2012). Integrated studies on the use of cognitive task analysis to capture surgical expertise for central venous catheter placement and open cricothyrotomy. *American Journal of Surgery*, *203*(1), 76–80.

Young, D., Harrison, D. A., Cuthbertson, B. H., & Rowan, K. (2013). Effect of early vs late tracheostomy placement on survival in patients receiving mechanical ventilation: the TracMan randomized trial. *JAMA*, *309*(20), 2121–2129.

Young, J. Q., Van Merrienboer, J., Durning, S., & Ten Cate, O. (2014). Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. *Medical Teacher*, *36*(5), 371–384.

Yujian, L., & Bo, L. (2007). A Normalized Levenshtein Distance Metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(6), 1091–1095.

Zendejas, B., Cook, D. A., Bingener, J., Huebner, M., Dunn, W. F., Sarr, M. G., & Farley, D. R. (2011). Simulation-Based Mastery Learning Improves Patient Outcomes in Laparoscopic Inguinal Hernia Repair: A Randomized Controlled Trial. *Annals of*

Surgery, *254*(3).

Zendejas, B., Wang, A. T., Brydges, R., Hamstra, S. J., & Cook, D. A. (2013). Cost: the missing outcome in simulation-based medical education research: a systematic review. *Surgery*, *153*(2), 160–176.