Advancing Inertial Measurement Unit Technology for Human Biomechanics and Engineering Education

by

Rachel V. Vitali

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mechanical Engineering)
in The University of Michigan
2019

Doctoral Committee:

Professor Cindy Finelli, Co-Chair
Professor Noel Perkins, Co-Chair
Professor James Ashton-Miller
Assistant Professor Ryan McGinnis, University of Vermont
Professor Kathleen Sienko

Rachel V. Vitali

vitalir@umich.edu

ORCID iD: 0000-0002-1436-6148

For my family and all the people who believed in me,
especially when I didn't.

ACKNOWLEDGEMENTS

I cannot fully or properly express my gratitude to everyone who has supported me throughout my doctoral program. It has been a journey these past four years, one that's had plenty of ups and downs. While this document largely reports on my successes, I have definitely had my fair share of failures as well. I am so appreciative of all the people in my life who knowingly or unknowingly helped me recover and persevere through those failures. As I write these acknowledgements, I recognize the inadequacy of my words and the limited space I have to thank only a few of the many who should be included.

First and foremost, my primary advisor, Noel Perkins, whose unwavering support and belief in me cannot be overstated. Six years ago, Noel took a chance on an admittedly overeager undergraduate in his introductory dynamics class (the same one at the center of my engineering education research). Since then, he has graciously and enthusiastically invested so much time and effort in cultivating whatever it was he saw in me back then. Who I am as a researcher, educator, colleague, and friend has been molded by his example as a truly remarkable person. I have thoroughly enjoyed working with you and never pass up an opportunity to tell others as much.

Next, my co-advisor, Cindy Finelli, who was absolutely critical in weaving together the two very distinct halves of my dissertation. I so admire your tenacity and resolve to bring engineering education research to the diverse academic landscape here at Michigan, and I am grateful to have been a part of it over the past few years. You integrated me into a new discipline and community that will undoubtedly shape me and my career in the years to come. Thank you for everything that you have done as an advisor, mentor, and friend.

I must also thank the rest of my phenomenal dissertation committee who have been so supportive of my somewhat unusual collection of work. James Ashton-Miller and Kathleen Sienko from the University of Michigan and Ryan McGinnis from the University of Vermont have all supported and shaped both halves of my work. I am so grateful for their insights, guidance, and feedback that improved my research.

Next, I thank the organizations that have supported me over the years, namely the the US Army Contracting Command-APG (Natick Contracting Division), the National Science Foundation, and the American Society of Mechanical Engineers. More specifically, I would like to thank Megan Coyne and Lee Hancock for their feedback and support for my current research as well as my future career. I would like to express my gratitude to the Mechanical Engineering Department and the Mechanical Engineering Graduate Council for providing an academic home for me for the past seven years, first as an undergraduate and then as a graduate student. Finally, I extend my sincerest appreciation of all of the people at the Center For Statistical Consultation and Research (CSCAR) who have provided excellent guidance for the statistical analyses I have conducted over the years.

All of the work presented herein has necessarily required the help of others. As an undergraduate learning about IMUs, I could not have asked for better labmates than Ryan McGinnis and Stephen Cain. Not only were they pivotal in helping me develop an understanding of the technology, they have since gone above and beyond in their support of me as mentors and friends. I look forward to continuing to work with both of you in the future. I also thank Leia Stirling, Antonia Zaferiou, Steven Davidson, Michael Potter, Neda Maghsoodi, Alyssa Mendoza, and Lauro Ojeda for their support and insights over the years as well. Andrew Lapointe, Jay Kim, Luis Nolasco, Deanna Gates, and Steve Broglio in particular supported the optical motion capture data collections and analyses for Chapter 4. I thank all of the undergraduate researcher assistants who assisted in the data collections for Chapter 5. Finally, I thank all of the participants who volunteered to be a part of the data collections. It could not have been possible without you.

For the engineering education research part of my life, Aaron Johnson and Jessica Swenson have both provided invaluable insight, mentorship, and friendship that

I am eternally grateful for and look forward to continuing in the future. I also thank Brian Self, Lisa Lattuca, Mark Guzdial, Caroline Crockett, Sarah Bork, Matt DeMonbrun, Trevion Henderson, Laura Carroll, and the rest of the EER community at Michigan for their support and insights that significantly improved my research. Thank you to all of the ME240 professors, GSIs, and undergraduate students who participated in the iNewton project. Your participation, engagement, and feedback is truly appreciated.

There is also a horde of others I need to thank. While I don't have the space to describe how each has played a critical role at various times in my life, I would at least like to mention some of them by name: Stephanie Pierce, Forrest Coghill, Nicole Liu, Samuelina Wright, Jeffrey Carpenter, Alison Hake, Callan Luetkemeyer, Emma Treadway, Deema Totah, Kim Ingraham, Maya Nath, Sara Troutman, Katherine Flanigan, Marie Rice, Ari Sasmal, James Gorman, Ashley Bielinski, Kim Novak, Rachel Casanova, John Laidlaw, Andy Poli, David Dowling, Jessy Grizzle, David Remy, Maani Ghaffari Jadidi, Jeffrey Scruggs, Adriene Beltz, Kristine Rezai, Elizabeth Batman, Emily Weitzman, and Kim Semmelroth. To the rest who are not on this list, I hope someday to thank you in person.

Last but certainly not least, I cannot thank my family enough. They have been so patient, understanding, supportive, and excited as I have immersed myself in my doctoral work. Throughout my life, my parents afforded me every opportunity they could that might help me be successful, and I, in turn, have tried to seize on every one I could. All three of you instilled in me the importance of self-sufficiency, consideration, ethics, enthusiasm, and so much more that shaped the person I am today, and for that I will forever be grateful. Likewise, my ridiculous, wonderful, amazing sister and brother are no less responsible for who I am today. I cannot thank you both enough for being the best siblings I could have possibly hoped for, and I look forward to a lifetime of shenanigans.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

ABSTRACT

Inertial measurement units (IMUs) are a ubiquitous technology found in navigation systems, mobile devices, and multiple products related to the Internet of Things. In its simplest form, an IMU contains a triaxial accelerometer and a triaxial angular rate gyroscope needed to deduce the six degrees of freedom of a rigid body. While their history traces back to navigation systems of aircraft, spacecraft, and satellites, IMUs now support new and innovative applications made possible through miniaturization via microelectromechanical systems (MEMS) fabrication methods. This thesis specifically advances the use of IMU technology within two important fields, namely, 1) human biomechanics and 2) engineering education.

Within the field of human biomechanics, this thesis makes two major contributions for using IMUs to quantify and understand human performance. The first deploys a pair of thigh- and shank-mounted IMUs to estimate the three-dimensional rotations across the human knee. This significant challenge requires a sequence of estimations that define: 1) the orientation of the IMU frames relative to their independent world frames, 2) the orientation of their independent world frames relative to each other, and 3) the orientation of the IMU frames relative to their respective body segment anatomical frames. Importantly, this thesis contributes a measurement theory to correct for the inevitable integration drift error arising in this sequence of estimates without reliance on magnetometer data. The theory exploits an anatomical kinematic constraint that the knee acts (predominantly) as a hinge. The resulting theory is first validated against data from high precision optical encoders embedded within a mechanical linkage and yields RMS differences of less than 5 degrees. The theory is further validated against data from conventional optical motion capture on human subjects (and across increasingly dynamic tasks) and yields overall RMS differences of less than 5 degrees. The second contribution leverages thigh- and upper

arm-mounted IMUs to define novel metrics of human crawling performance and technique to support the evaluation of warfighters. Performance metrics derived from the raw IMU data successfully distinguish superior from inferior crawling performance and the degradations in performance from added body-borne loads.

Within the field of engineering education research, this thesis contributes a thorough investigation of an active learning intervention that employs IMUs to explore concepts in an introductory engineering dynamic course (ME240 at the University of Michigan). The intervention takes three forms that elicit increasing cognitive engagement per Chi's ICAP framework, namely: 1) Demonstrations, 2) Prescribed Experiments, and 3) Student Projects. Building from a foundation of supporting literature and learning theories, this research tests the hypothesis that students who engage with the active learning IMU intervention will demonstrate positive responses in 1) conceptual understanding, 2) self-efficacy, and 3) intention to persist relative to students who do not (control). As measured solely by the Dynamics Concept Inventory, the active learning IMU intervention elicits little change in conceptual understanding relative to the control. By contrast, as measured by a modified version of the Longitudinal Assessment of Engineering Self-Efficacy, the active learning IMU intervention elicits significantly higher course-specific self-efficacy and intention to persist in the field relative to the control.

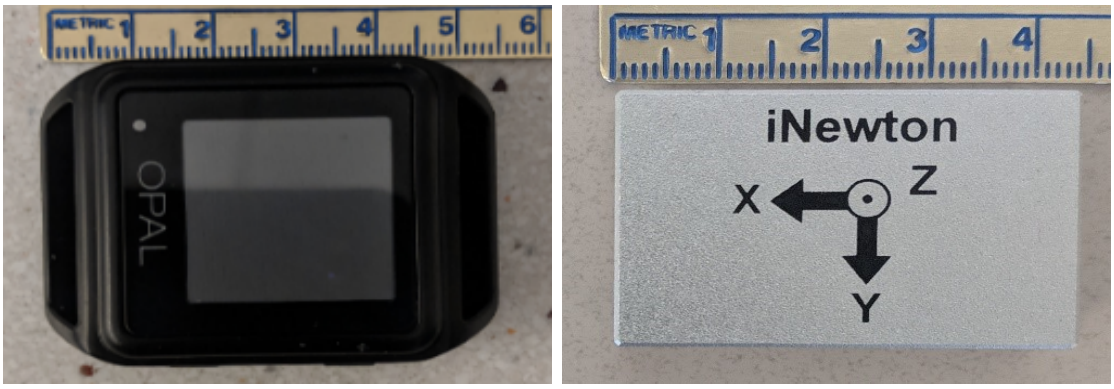
CHAPTER 1

Introduction and Overview

1.1 Opportunities for Advancing IMU Technology

Inertial measurement units (IMUs) are a ubiquitous technology found in navigation systems, most mobile devices, and a growing number of products in the space of the Internet of Things (IoT). The history of IMUs traces back to their core function within inertial navigation systems to support the navigation of aircraft, spacecraft, and satellites (e.g. [1, 2]). With appropriate signal processing techniques, the data harvested from IMUs are used to resolve the movements of a rigid body to which they are affixed (e.g. orientation of spacecraft [3, 4]). Outside of their original aerospace applications, IMUs are finding new and extraordinary implementations in the wake of their miniaturization by microelectromechanical systems (MEMS) fabrication methods. These fabrication methods have created a class of IMUs often referred to as tactical grade devices, which is in contrast to the highly accurate, often much larger inertial grade devices required for precision navigation [5]. MEMS IMUs are inexpensive and suitable for many consumer products and applications [6]. Among the wide and varied applications of MEMS IMUs are those used in robotics, ground vehicles (including personal transportation systems such as the Segway™), gaming devices, virtual and augmented reality technology, mobile devices, personal computers, monitoring devices for civil structures, sports equipment, and other consumer products. Adding to this broad context of applications, this thesis advances

the use of IMU technology within two important fields: human biomechanics and engineering education. The figures below show the two versions of IMUs employed in this work, respectively.



(a) (b)

Figure 1.1: Two versions of inertial measurement units. **(a)** APDM Opal. **(b)** iNewton.

In its simplest configuration, an IMU contains a triaxial accelerometer and a triaxial angular rate gyroscope, both of which produce signals sampled simultaneously by a microprocessor. Additional sensors like magnetometers, barometers, and global positioning systems (GPS) are sometimes included in the design, and the supplemental data are used to improve estimates of orientation and position through a variety of sensor fusion techniques [7]. When coupled with radio technology, individual IMUs can form a multi-sensor system providing synchronized data to more easily resolve the coordinated motions of multi-body systems, as in the analysis of human motion described further below. The overall improvements over past decades in MEMS fabrication methods have produced highly miniaturized, low-power, and inexpensive IMUs to support exciting new applications, including the novel applications to human biomechanics and engineering education that are the focus of this dissertation.

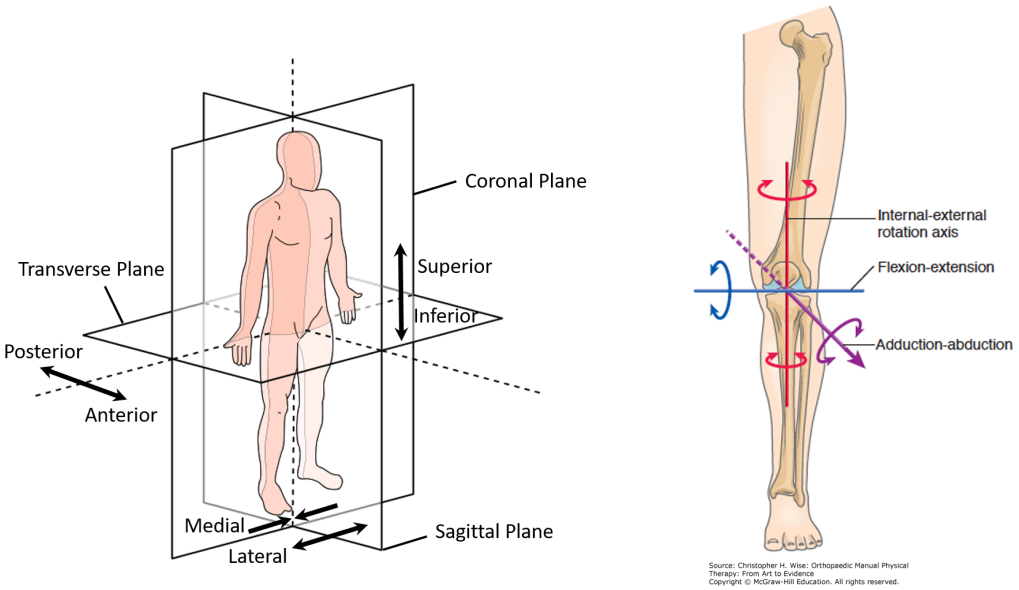
1.2 Part I: Human Biomechanics

1.2.1 Motivation and Potential Impact

In recent years, IMUs have gained considerable popularity in human biomechanics investigations. As a motion capture technology, IMUs present an attractive alternative to more traditional methods because of their versatility, portability, and relatively low cost [8]. Position-based methods like optical motion capture (MOCAP) restrict movements to a finite (and often modest sized) capture volume, rendering this method difficult to use in naturalistic environments (see, for example, [9, 10]). In addition, MOCAP requires considerable set-up, post-processing time, and training. IMU-based motion capture does not suffer from these limitations, particularly when post-processing is automated in specialized mobile applications. However, IMU-based motion capture introduces other challenges, most notably sensor integration drift error that limits the accuracy of IMU-derived orientation, velocity, and position estimates [11]. Despite the challenges, there are profound implications of deploying wearable (and inexpensive) IMU technology for measuring human biomechanics in natural contexts. For example, consider the potential impact of this wearable technology for applications in human health, worker safety, athletic performance, warfighter performance, among many other applications.

Among the many possible applications of body-worn IMUs in biomechanics, the first part of this thesis contains two distinct objectives. A common aim of many IMU-based human biomechanics studies is to recreate more traditional performance metrics (e.g. joint angles) derived from MOCAP-based studies. As such, the first objective focuses on the kinematics of the human knee, specifically three-dimensional (3D) rotations across the knee joint. There is strong motivation in doing so. The kinematics of knee rotations have significant ramifications for human health and performance and in a variety of contexts including human mobility, worker safety and health, athletic performance, and warfighter performance. At first glance, the knee joint may appear to function simply as a hinge (one dimensional rotation) undergoing flexion-extension. In reality, it is considerably more complex that allows 3D rotations

namely flexion-extension, internal-external rotation, and abduction-adduction as is illustrated in Fig. 1.2b. Annually, approximately 2.5 million sportsrelated injuries occur in adolescent athletes [12]. Furthermore, recent evidence suggests that ACL injuries may actually be overuse injuries [13], in which case measuring the number and severity of loading cycles each week may prove useful in athletes of any age. The knee is also one of the most common joints for extensive reconstruction or replacement, both of which require committed rehabilitation crucial to the healing process for a successful outcome (e.g., [14]), for which there is significant motivation for non-invasively measuring/monitoring and understanding the 3D rotations across the human knee for long periods of time outside of a laboratory. Furthermore, with the advent of MEMS IMUs, there is now an opportunity to accomplish this using an inexpensive and wearable technology.



(a) (b)

Figure 1.2: **(a)** Anatomical planes and directions referred to in the document. **(b)** Anatomical rotations of the knee. Image adapted from [15].

By contrast, great strides have also been taken to tap into the potential inherent to this technology by defining new (nontraditional) metrics for well-defined tasks that

draw on the strengths of this technology. Thus, the second objective of the human biomechanics portion of the thesis focuses on defining and evaluating human performance in the context of an outdoor obstacle course used by military organizations worldwide to quantify the effects of clothing and individual equipment (CIE). This obstacle course, introduced by the US Marine Corp as the Load Effects Assessment Program (LEAP), incorporates a standard (10-task) obstacle course that embeds combat-relevant movements and tasks [16]. Physical performance in the LEAP obstacle course is historically quantified by the overall time to complete the course together with the times to complete each obstacle (measured using timing gates). Although the times to complete an entire course or any obstacle within it are valid measures of performance, the times alone do not reveal the underlying biomechanical movements that further discriminate performance levels or reveal sub-movements that limit or enhance performance. Of course, measuring human movement in the context of a large outdoor obstacle course ($\sim 1800\ m^2$ in the context of this work) is largely precluded using standard MOCAP methods, whereas IMUs offer an alternative to address this challenge. The specific task under investigation in this dissertation is the high crawl, in which participants crawl on elbows and knees as quickly as possible [17].

To this end, the overall objective for the first part of this dissertation is to advance the use of wearable IMU technology for accurately estimating the 3D rotations across the human knee joint and for quantifying and evaluating human crawling performance in an outdoor obstacle course. These objectives are achieved by completing three major tasks, namely:

1.1 Demonstrate a proof of concept for deterministically aligning world frames from independent sensors for a measurement theory for estimating 3D knee rotations

1.2 Develop and validate a purely probabilistic measurement theory for estimating 3D knee rotations

1.3 Validate the measurement theory on human subjects with MOCAP

1.4 Develop and evaluate human crawling performance metrics

A brief summary for each of these major tasks is provided next, and each section represents a chapter of this dissertation. The first portion of this dissertation provides a validated measurement theory for deploying wearable IMUs for resolving 3D knee rotations, which will also provide a foundation for extending the method to other major skeletal joints. Simultaneously, a thorough investigation of human crawling performance demonstrating the power of nontraditional performance metrics in discriminating superior crawling performance as well as quantifying the effects of load on performance.

1.2.2 Chapter 2: Measurement Theory for Estimating Knee Rotations with Deterministic World Frame Alignment

Task 1.1 is accomplished in Chapter 2, which contributes a theory of estimating 3D rotations across the knee via IMUs attached to the thigh (femur) and shank (tibia). Rotation across the knee is defined as the relative orientation between the femur and shank, which requires knowledge of the orientation of each bones anatomical frame of reference. This requirement creates two major challenges for IMU-based methods. First, IMU-based methods are fundamentally incapable of adhering to the International Society of Biomechanics (ISB) convention [18] for defining anatomically significant frames of reference for each bone, which relies on locating specific bony anatomical landmarks. It is impractical to resolve the location and orientation of the IMUs relative to these landmarks and so that information is essentially unavailable with IMU-based methods. Second, each IMU is an independent measuring device and so one cannot estimate the orientation of one IMU (e.g. one attached to the thigh) relative to a second IMU (e.g. one attached to the shank) without first knowing the orientation of each IMU relative to a common world frame of reference. Being independent devices, the two IMUs do not create identical world frames, even if they are coupled with magnetometers (i.e., unequal magnetic field distortion from ferromagnetic materials). This chapter lays out the methods for resolving both of these challenges starting with an idealized version of a knee.

The first challenge (deducing anatomically-relevant directions for the knee) is addressed by introducing simple functional alignment movements that allow one to estimate the medial-lateral axis of the knee and the superior-inferior axes of the tibia and femur. The orientation of the shank and thigh are then deduced relative to these "subject"-calibrated axes and from this follows anatomically-relevant estimates of 3D rotations across the subjects knee. The second challenge is addressed by exploiting a kinematic constraint [19] to create a common world frame from the independent world frames of each IMU. The constraint follows from identifying time intervals when the knee is functioning largely as a hinge joint and using the orientation of the hinge axis to essentially "align" the two independent world frames of the IMUs. When used in tandem as described in this chapter, the two methods yield a measurement theory that successfully resolve 3D rotations across a joint. The validity of these methods, particularly the proof of concept for the kinematic constraint, is assessed using ground truth data collected from highly precise optical encoders embedded in the rotational joints of a coordinate measurement machine (CMM). Two IMUs are attached to links on the CMM (analogous to the shank and thigh) such that there are three degrees of rotational freedom between them to mimic a human knee joint. The resulting measurements clearly demonstrate that the theory can indeed resolve 3D rotations across a mechanical joint with high precision (e.g., RMS errors less than 4°) [20]. In light of this success, albeit with an idealized mechanical testing apparatus, an additional study was conducted to validate and extend the measurement theory as described next.

1.2.3 Chapter 3: Measurement Theory for Estimating Knee Rotations with Probabilistic World Frame Alignment

Chapter 2 addresses Task 1.2 by further developing the theory of estimating 3D rotations across the knee by including estimates of IMU orientation from a robust Error-State Kalman Filter (ESKF) and replacing the deterministic treatment of the kinematic constraint with a probabilistic treatment in the ESKF formulation. This chapter begins with a brief overview of approaches used for state estimation with

an emphasis on those under the broad Kalman filtering umbrella. The first study seeks to implement a robust EKSF with a novel treatment of magnetometer data when local magnetic interference is detected. Specifically, when the IMU is static and the magnetometer data is largely constant, this data is still provides a constant direction of the magnetic field that can be used to correct integration drift error about vertical. As a result, the agreement between the IMU-estimated orientation and the orientation provided by the CMM is excellent (e.g., RMS errors less than 5° over a 5-minute period).

The second study then seeks to extend the robust ESKF from a single IMU to two IMUs to then estimate the 3D angles between their respective measurement frames. This is accomplished by incorporating the kinematic constraint described in Chapter 2 in which the knee analog frequently acts as a hinge. Specifically, the deterministic world frame correction from Chapter 2 is incorporated into a measurement model to be used to update the orientation of one of the sensors relative to the other. While the results in Study 1 demonstrated a strong dependence on magnetometer data for accurate estimates of *absolute* orientation for a single IMU, this study compares the accuracy of the *relative* orientation estimates between two IMUs and with less dependency on magnetometer data. Specifically, the kinematic constraint and the novel magnetometer treatment are used as strategies to correct the integration drift error between the sensors. Furthermore, a method is introduced for estimating the orientation of each sensor relative to its respective link frame (i.e., its body-fixed frame) without relying upon the alignment movements used in Study 1 and Chapter 2. Doing so reduces yet another source of error and allows for a clearer interpretation of the efficacy of the ESKF. As a result, the agreement between the IMU-estimated orientation and the orientation provided by the CMM is excellent (e.g., RMS errors less than 6° over 2-minutes of challenging movements). Moreover, the probabilistic treatment of the kinematic constraint is revealed to be far superior to the deterministic treatment from Chapter 2. Thus, an additional study is conducted to validate the measurement theory developed in this chapter on human subjects and with independent measurements provided by optical motion capture (MOCAP).

1.2.4 Chapter 4: Application of Measurement Theory to Human Subjects and Validation with MOCAP

Task 1.3 is accomplished in Chapter 4, which describes the application and validation of the measurement theory using human subjects and with independent measurements from optical motion capture (MOCAP). In the preceding chapter, complexities of the human knee were largely absent due to the rigidity and unambiguity of the rotational joints of the CMM. By contrast, present in experiments on human subjects are the difficulties arising from imprecise determination of the knee anatomical axes, laxity of the knee, superimposed motions of the underlying soft tissues, among others. To establish the validity of the measurement theory in the presence of these complexities, human subjects complete a range of increasingly dynamic tasks to determine when and how rotation estimates from the IMU-based measurement theory differ from those provided by MOCAP. The results of this investigation have two important implications. The first is a demonstration that the anatomical kinematic constraint that the human knee acts predominantly as a hinge is reliable in correcting relative integration drift error about vertical. The second reveals that estimation errors due to skin artifact manifest differently in the two motion capture modalities. Furthermore, the method introduced in Chapter 3 for determining the relationships between the IMU body-fixed frames and the CMM's links is implemented here to determine the relationship between the IMU frames and their respective "true" anatomical frames.

Finally, one of the significant, open challenges associated with IMU-based estimation of joint angles relates to how the anatomical frames are defined. This chapter also includes a systematic survey of the different methods reported in the literature ($n = 112$) for defining anatomical frames of reference for inertial motion capture, which are roughly categorized into one of four approaches. The first and most common method is functional alignment movements, which consist of a human subject completing one or a set of movements from which at least one anatomical axis can be estimated. The other most common method is to assume the sensor orientation is aligned with the body segment's anatomical frame, which is accomplished by the re-

searcher carefully attaching the IMU to the body segment. Within the last few years, self-aligning methods have been proposed in which the body segment anatomical axes are estimated by exploiting a kinematic model or assumption of the joint. Finally, some researchers have proposed using external information from a source other than the IMUs to determine the relationship between the IMU body-fixed frame of reference and the anatomical frames. There are significant variations within each method thereby highlighting the fact that there does not exist a common convention for defining anatomical frames of reference for inertial motion capture despite the exponential growth in studies utilizing IMUs for human biomechanics research.

1.2.5 Chapter 5: Human Crawling Performance Metric Development and Evaluation

Task 1.4 is addressed in Chapter 5, which includes developing and evaluating metrics of performance for a crawling task defined in the context of an outdoor obstacle course. For military organizations worldwide, crawling represents a type of locomotion that is still operationally relevant to this day. Crawling has implications for assessing functional capacity following injury [21], exposing the mechanisms and types of injury [22, 23], and optimizing the size and configuration of carried loads [24–28]. Historically, crawling performance for military assessment is quantified solely by the time required to complete the task. To that end, four performance metrics were developed to describe and distinguish crawling performance in population of human subjects. Crawl speed is defined as the average speed the crawls at to complete the task. Crawl stride time is the mean of the stride times defined by elbow strikes detected in the data collected by upper arm-mounted IMUs. Two coordination metrics describe the phasing of the ipsilateral (same side) and contralateral (opposite side) pairs of upper arms and thighs as determined from data collected at the upper arms and thighs.

A thorough statistical analysis of these metrics reveal these four metrics can effectively distinguish superior crawling performance. In particular, faster crawl speeds (denoting superior performance) are accompanied by shorter crawl stride

times, largely in-phase contralateral limbs, and largely out-of-phase ipsilateral limbs. In light of this success, an additional study was conducted to investigate the relationship between these performance metrics and body-borne loads. In previously conducted studies [24–28], the results suggest that added load deteriorates performance as defined by a timing metric. However, the mechanisms driving that degradation have largely been unknown until now. Additional human subjects were recruited to completed the same crawling task four times: twice unloaded, once carrying 15% of their body weight, and once carrying 30% of their body weight. The goal of this study was to quantify and characterize the changes in performance due to the additional load for which the hypothesis was guided by the results from the previous study. In particular, the results confirm that performance degrades such that slower crawl speeds are accompanied by longer crawl stride times and less coordinated limbs.

1.3 Part II: Engaged Learning of Engineering Dynamics

1.3.1 Motivation and Potential Impact

The second part of the dissertation leverages IMU technology (called iNewton) as a platform for active learning of Newtonian mechanics to support undergraduate education in engineering. Active learning is broadly defined as any instructional practice that involves students in the learning process, and common approaches include techniques like cooperative learning, problem-based learning, and experiential learning [29]. A systematic review conducted by Prince [30] provides strong evidence that active learning results in positive gains in student learning and broader educational experiences across a wide variety of science, technology, engineering, and mathematics (STEM) fields. This mode of learning is largely grounded in constructivist learning theory, which posits that students construct new knowledge by connecting new ideas and experiences to existing ideas and experiences to form new or enhanced understanding [31]. John Dewey, arguably the philosophical founder of this learning theory, says, "*To 'learn from experience' is to make a backward and forward connection between what we do to things and what we enjoy or suffer from*

things in consequence. Under such conditions, doing becomes a trying; an experiment with the world to find out what it is like" [32, p. 164]. As this applies to learning Newtonian mechanics, a logical hypothesis is providing students with authentic opportunities to connect the mathematical models describing the physical phenomena will improve their conceptual understanding of engineering dynamics. Additionally, Betz and Schifano [33] demonstrated that active learning can improve self-efficacy (the strength of one's belief in knowledge and skills needed to achieve a task) while Ohland et al. [34] showed that academic engagement via active learning is a precursor to persistence (and by association intention to persist [35]).

Given the ease with which IMUs yield motion (kinematic) data, they provide a novel and ready-made platform to provide authentic opportunities to explore and learn Newtonian mechanics as a new form of engaged learning. Consequently, this part of the dissertation represents a focused study within the field of Engineering Education Research (EER) to assess how best to provide these opportunities in such a way that are beneficial for students while also considering the feasibility for instructors to implement this in the future. Through this study, IMU technology is systematically introduced as a new learning intervention to an otherwise traditional lecture-based classroom for a large undergraduate course at the University of Michigan. The learning intervention takes the form of simple experiments (with varying levels of student engagement) that exploit IMU data to reveal important and commonly misunderstood concepts in particle and rigid body dynamics. Importantly, these studies endeavored to quantify and compare the effects of the different types of the intervention on student conceptual understanding, self-efficacy, and intention to persist as summarized further below.

The overall objective for the second part of this dissertation is to advance the use of IMU technology for engaged learning of Newtonian mechanics for engineering education. This will be achieved by systematically scaling up the use of IMUs in the classroom, which includes Demonstrations, Prescribed Experiments, and Student Projects, and measuring the effects of the intervention on the students. Thus, this objective is achieved by two main research tasks, namely:

2.1 Quantifying and evaluating the cognitive effects of this intervention (i.e., conceptual understanding of engineering dynamics)

2.2 Quantifying and evaluating the non-cognitive effects of this intervention (i.e., self-efficacy and intention to persist)

Summaries of the major tasks is provided as well as descriptions of the theoretical basis for this research and the study design.

1.3.2 Chapter 7: Active Learning in Undergraduate Engineering

Background, a conceptual framework, and the learning theories that guide the study and analysis are presented in Chapter 7. From the literature, research across many disciplines has shown that active learning techniques positively affect conceptual understanding, self-efficacy, and intention to persist. A conceptual framework relating these cognitive and non-cognitive effects to each other as well as to active learning is proposed. Finally, two theories provide a basis for the study's hypothesis and the conceptual framework. The first learning theory guiding this work is constuctivism, which theorizes that students construct their own understanding of new concepts through a process that depends on their prior knowledge and experiences [36]. The other is Albert Bandura's psychological theory of self-efficacy, which *"refers to beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments"* [37, p. 3]. Constructivism and self-efficacy provide a basis for why the different levels of the intervention are expected to positively impact conceptual understanding, self-efficacy, and intention to persist.

1.3.3 Chapter 8: iNewton Study Design

This chapter provides a detailed description of the study design including the class setting, a pilot study, participant demographics for the current study, and descriptions for each active learning IMU intervention level. The study was designed such that each subsequent intervention level increases the cognitive engagement the students have with the IMU technology, which can be described by Chi's ICAP

framework [38] (also described in Chapter 7). This framework categorizes students' engagement activities based on students overt (observable) engagement behaviors, which correspond to one of four modes: *Interactive*, *Constructive*, *Active*, and *Passive* [39]. The first level of this intervention takes the form of two Demonstrations conducted in class for the students (*Passive*). The second intervention level takes the form of two Prescribed Experiments the students conduct on their own outside of class with IMUs provided to them (*Active*). The third and final level of this intervention takes the form of Student Projects for which the students design and conduct experiments of their own imagining that are meant to reveal some subset of course concepts (*Constructive*).

1.3.4 Chapter 9: iNewton Cognitive Effects

Chapter 9 addresses Task 2.1, which contains several subtasks in addition to evaluating the cognitive effects of the active learning IMU intervention. Monfort, Brown, and Pollock state, *"Conceptual understanding (peoples personal explanations of how and why the world works) is knowledge in context, and is therefore more transferable than computational ability"* [40, p. 111]. For this work, student conceptual understanding is measured using a concept inventory called the Dynamics Concept Inventory (DCI) [41, 42]. Concept inventories, like the DCI, are frequently used to assess student understanding of a specific set of concepts, typically in a low-stakes setting. These concept inventories are typically evaluated for their validity and reliability, which is frequently repeated in a specific study's context.

Thus, the first subtask (study) investigates how student performance during a low-stakes evaluation of conceptual understanding relates to performance during a high stakes evaluation. The synergy of the relationships between these evaluations indicate the low-stakes overall DCI assessment is an authentic measure for quantifying student conceptual understanding. Spurred by the results presented by Jorion et al. [43], the second study evaluates the validity and reliability of the individual items on the DCI with the goal of eliminating less informative items. As a result of these analyses, four items were eliminated from the analysis for the final subtask

evaluating the cognitive effects of the IMU-based active learning intervention. The results of this analysis reveals the intervention had limited impact on conceptual understanding as measured by the DCI. However, the evaluation of the DCI for the second subtask and additional analyses conducted in the final subtask reveals the DCI in its current form lacks the ability to differentiate conceptual understanding of engineering dynamics concepts on a finer grain.

1.3.5 Chapter 10: iNewton Non-Cognitive Effects

The second task (Chapter 10) addresses quantifying the non-cognitive effects that this intervention may have on students' self-efficacy and intention to persist. Self-efficacy is a term that, in this context, focuses on the student's beliefs that he or she will be successful in engineering in general as well as in this specific introductory engineering dynamics course [44, 45]. As previously mentioned, both conceptual understanding and self-efficacy have ramifications for students' intention to persist, which is defined as students' choice to continue with the major. The final non-cognitive response investigated in this chapter is student affect, which relates to how students feel about engaging with the different intervention levels. Thus, these four responses (Engineering Self-Efficacy, Course-Specific Self-Efficacy, Intention to Persist, and Student Affect) are measured with a modified version of an instrument called the Longitudinal Assessment of Engineering Self-Efficacy (LAESE) [46]. The first study evaluates the validity and reliability of the four constructs in the modified LAESE before proceeding to evaluating the effects of the active learning IMU intervention on each of them. The next three studies leverage the findings from the first by explicitly evaluating the effects of the intervention on self-efficacy, intention to persist, and student affect, respectively.

This first study confirms the validity and reliability of the four constructs. A second study examines the effects of the active learning IMU intervention on Engineering Self-Efficacy (ESE) and Course-Specific Self-Efficacy (CSSE) and exhibits mixed results. Decreases in ESE are associated with increasing levels of the intervention (i.e., increasing levels of cognitive engagement). However, the significant

interaction between ESE Pre-score and intervention level indicates that students who enter the class with high ESE are also relatively unaffected by the intervention, which could be evidence of students who have developed a resilient sense of self-efficacy. For CSSE, there is a significant increase for students who engaged with the Student Projects version of the intervention. A third study reveals differences in Intention to Persist (PER) are significantly associated with the active learning IMU intervention. The largest increases in PER, particularly for students with lower PER at the start of the semester, relate to greatest engagement with the IMU technology. Specifically, Prescribed Experiments and Student Projects are associated with statistically significant increases in PER with the gains associated with Student Projects being the largest. A final (fourth) study confirms that students have a positive Student Affect (SA) towards the interventions. While there is no significant difference in SA between Demonstrations and Prescribed Experiments, there is a significant increase in SA associated with Student Projects.

Part I

Human Biomechanics

CHAPTER 2

Measurement Theory for Estimating Knee Rotations with Deterministic World Frame Alignment

2.1 Introduction to Motion Capture

The human knee is susceptible to a number of injury mechanisms including non-contact hyperextension, over-use, and direct impact (see, for example, [47, 48]. These injuries can manifest as a number of different health problems like osteoarthritis, which often severely limit mobility and activities of daily living [49, 50]. Accordingly, the three-dimensional (3D) rotations across the knee (flexion-extension, internal-external rotation, and abduction-adduction) serve as important markers of knee health and performance, and in multiple contexts such as human mobility, worker safety and health, athletic performance, and warfighter performance. For example, clinicians assess knee laxity and rotations to diagnose knee injuries and to determine the need for medical interventions of varying degrees like physical therapy, injections, or knee arthroplasty [14, 51]. Many biomechanical analyses incorporate measurements of knee rotations including studies of the long-term effects from knee injuries [52, 53], joint disorders including arthritis [54], and age- and gender-related differences in knee health [55]. By analyzing knee rotations, researchers have also explored the effects of load carriage related to fall prevention [56], metabolic cost during

walking [57, 58], and warfighter performance during walking [59]. Warfighter performance in particular is difficult to study since the conditions under which warfighters perform cannot be recreated well in a laboratory, but restricting studies to a laboratory have other ramifications as well. For example, unconstrained walking gait over ground has been found to differ from gait employed while walking on a treadmill [60–62].

The several commonly used methods for measuring knee rotations, namely goniometers, magnetic motion capture, and optical motion capture, are largely restricted to laboratory settings. Uniaxial goniometers for measuring flexion-extension are either purely mechanical (akin to a protractor) or electromechanical (e.g. potentiometers), as are multiaxial goniometers for measuring any combination of flexion-extension, internal-external rotation, and adduction-adduction [51, 54, 63]. In any case, errors are induced by soft tissue sensor mounting as well as misalignment of the goniometer sense axes relative to the knee anatomical axes. Magnetic motion capture triangulates the location of magnetic sensors rigidly attached to major body segments via a magnetic field generated by a transmitter, which restricts the capture volume to modest sizes [64]. Errors can arise from low sampling rates, latency, and magnetic field disturbances from electrical currents (i.e. wiring) and ferromagnetic materials (i.e. building materials).

Most commonly, knee rotations are deduced from optical motion capture data, which can be collected one of two ways. The more common method involves placing markers on anatomically prominent bony landmarks located by the investigator by palpating body segments. These markers are either passive or active. Passive markers are retroreflective and are triangulated by light-emitting infrared cameras placed around the capture space (see, for example, [14, 53, 56, 59, 65]. Active markers produce the light via LEDs to be triangulated by cameras, which requires a battery source and oftentimes the markers are daisy chained within a suit worn by the subject [66]. Because these markers are the light sources instead of reflecting the light from infrared cameras, the capture volumes are typically much larger with a higher signal-to-noise ratio [67]. The less common method is markerless and requires complex algorithms to resolve the motion of each body segment [68, 69]. Despite significant

advances in this methodology, the current state of these algorithms are not yet at the point that they can resolve body kinematics to a level of precision and accuracy comparable to the more traditional methods. Regardless, optical motion capture methods place limits on the capture volume and may also suffer from occasional marker occlusion, soft tissue movement relative to the underlying skeletal structure, and marker placement precision on approximate bony landmarks [67]. Although magnetic and optical motion capture share the same benefits of being position-based, they still tether studies to a laboratory setting.

Body-worn inertial measurement units (IMUs) provide an attractive, alternative means to estimate the 3D rotations across the knee from data collected from the on-board accelerometers, angular rate gyros, and magnetometers (if available). Unlike the above methods that are largely restricted to laboratory settings, body-worn IMUs may readily be used outside the laboratory, thereby potentially increasing the validity of research conclusions by enabling data collections in the real world. To estimate 3D rotations across the knee, one must first estimate the relative orientation of the anatomical axes of the thigh to those of the shank, a result that does not immediately follow from shank- and thigh-mounted IMUs as they are independent devices [20]. However, this result is obtainable if the orientation of each IMU is first established relative to a common world frame of reference, which is a technical challenge addressed in this chapter. Furthermore, results from IMU-based estimates will only agree with more traditional motion capture, like optical motion capture, if the anatomical frames of reference for each body segment can be identified, which is another challenge addressed in Chapter 4.

2.2 Inertial Motion Capture Literature Review

A number of prior studies propose strategies for estimating rotations across a variety of joints using body-worn IMUs. Luinge et al. [70] deduce the relative orientation of two IMUs by exploiting a wrist joint constraint. Similarly, Müller et al. [71] determine the rotation axes for and ultimately the angles across the elbow via an optimization algorithm built on the condition that the joint has only two degree

of freedom. Using the medial-lateral axis of the knee as an anatomical constraint, Cooper et al. [72] estimate (planar) knee flexion-extension during straight-line walking through fast running. Following suit, Seel et al. [19] also estimate knee flexion-extension after first constructing the knee medial-lateral axis using angular rate data from shank- and thigh-mounted IMUs assuming their world frames are identical. In addition, they fuse gyro- and accelerometer-based knee flexion-extension estimates to significantly reduce the effects of drift during long trials. An extension is offered by Laidig et al. [73] where knee flexion-extension angles are accurately estimated by exploiting the knees hinge axis to control misalignment about the vertical axis due to drift and/or magnetic field interference. By contrast,Favre et al. [74] exploit the anterior-posterior axis of the knee as an anatomical constraint to also estimate knee internal-external rotation and abduction-adduction (i.e., full 3D rotations across the knee) through an intermediate step employing the distinct world frames of the two IMUs. The misalignment between the distinct world frames is estimated using an assumed constant correction angle for relatively short duration trials. An extension is offered by Brennan et al. [75] through a time-varying correction angle using correction estimates at the start and the end of each trial and by requiring the IMUs return to their original orientations. This method is validated using an instrumented gimbal that provides ground truth data from embedded optical encoders.

While the above methods do not consider IMUs that include magnetometers, other methods do so and use magnetic North to align the IMU world frames as well to estimate (yaw) drift about the vertical axis [76, 77]. An overall approach to fusing magnetometer and inertial sensor data is outlined in [78] which also considers corrections for magnetic field interference. However, the estimates of magnetic North from two IMUs may differ due to discrepancies in magnetometer data [79, 80] despite these corrections, ultimately limiting this advantage. Reducing these discrepancies may follow from improving IMU hardware, updating filter parameters, or including additional (and complementary) sensors for fusion [81]. The remainder of this chapter will detail a new method of controlling orientation drift by exploiting an anatomical kinematic constraint specific to the knee, followed by results from a validation conducted with a coordinate measurement machine (CMM) (MicroScribe

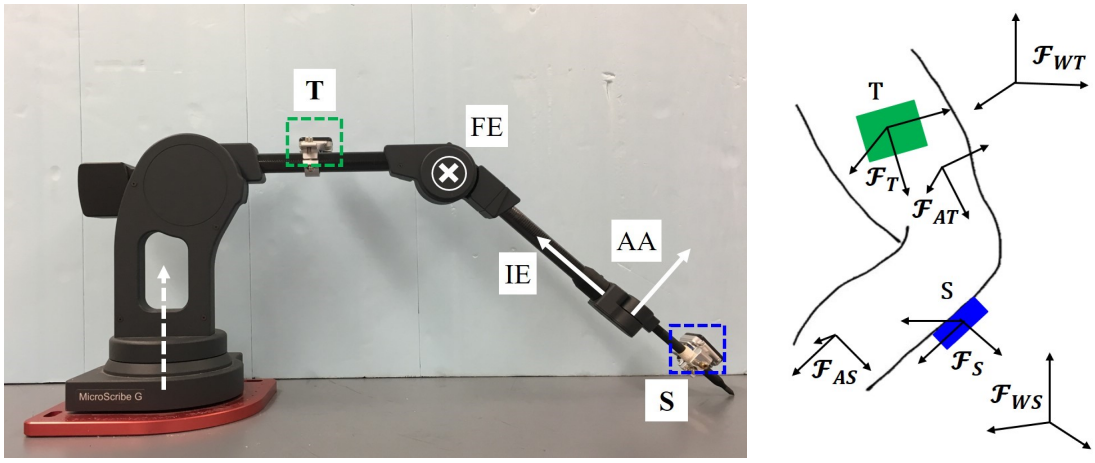
G2X, Solution Technologies, MD, USA).

2.3 Methods

The CMM show in Figure 2.1a embeds high precision optical encoders (0.0003° resolution [82]) that measure rotations about three axes representing knee flexion-extension (FE), internal-external rotation (IE), and abduction-adduction(AA). Two IMUs (Opal sensors, APDM, Portland, OR, USA; sensor characteristics and orientation estimate information available at `http://www.apdm.com/wearable-sensors/`), rigidly mounted to the illustrated two links replicate the functions of thigh- and shank-mounted IMUs (**T** and **S**, respectively). The resulting apparatus enables direct comparison of IMU-estimated knee FE, IE, and AA to measured values from the three high precision optical encoders and over a wide range of simulated knee movements.

2.3.1 Experimental Procedure

For the validation, data from the two IMUs are first time-synchronized to the encoder data from the CMM. The assembly is rotated by hand about the CMMs base (white axis in Figure 2.1a) with the three knee axes (FE, IE, AA) locked. The angle measured by the optical encoder about the base (dashed white) axis is differentiated with respect to time yielding an angular velocity signal to compare with those measured by the thigh (green) and shank (blue) IMUs. The data from the two IMUs are already time-synchronized, and their synchronization with the data from the CMM follows from measuring (and subsequently subtracting) the time delay between their respective angular rates. Next, the two functional alignment movements are conducted to estimate the anatomical axes of the shank and thigh for the knee analog. First, the superior-inferior axes of the shank and thigh are estimated by holding each segment still (for approximately 10 seconds) while vertical. The measured acceleration for each segment defines the direction of gravity, which is also aligned with the superior-inferior axis of each segment. Next, the medial-lateral axis



(a) (b)

Figure 2.1: Knee analog formed by a coordinate measurement machine (CMM). **(a)** Three anatomical axes for flexion-extension (FE), internal-external rotation (IE), and abduction-adduction (AA) are labeled are the corresponding rotational joints of the CMM. Two labeled IMUs are mounted to the CMM with **T** (green) analogous to a thigh-mounted IMU and **S** (blue) analogous to a shank-mounted IMU. **(b)** Definitions of three frames of reference for a human knee associated with a shank-mounted IMU (blue) including the shank IMU frame, $\mathcal{F}_{\mathbf{S}}$, the shank anatomical frame, $\mathcal{F}_{\mathbf{AS}}$, and the shank IMU's world frame, $\mathcal{F}_{\mathbf{WS}}$. Analogous frames of reference are illustrated for the thigh-mounted IMU (green).

is established using essentially the procedure outlined in [19]. In particular, the CMM is exercised purely about the flexion-extension axis with the two remaining knee axes locked. In so doing, the knee acts as a pure hinge, and one can readily compute the medial-lateral (hinge) axis with respect to the sense axes of each IMU. The resulting medial-lateral axes, so measured by the thigh- and shank-mounted IMUs, play a key role in the estimation process described below. Finally, four characteristically distinct knee movements for generating the truth data are considered for validation. These movements, each repeated for $N = 50$ trials, include: 1) pure flexion-extension, 2) pure internal-external rotation, 3) pure abduction-adduction, and 4) combinations of all three rotations. Each type of movement is made by hand, and one trial lasts approximately 10 seconds (with the appropriate CMM axes either free or locked).

2.3.2 Defining Segment Frames of Reference

Figure 2.1b illustrates three distinct frames of reference associated with each of the thigh- and shank-mounted IMUs. In particular, note the shank IMU frame, $\mathcal{F}_{\mathbf{S}}$ (defined by the IMU sense axes), the shank anatomical frame, $\mathcal{F}_{\mathbf{AS}}$, and the shank IMU world frame, $\mathcal{F}_{\mathbf{WS}}$. The three analogous frames of reference ($\mathcal{F}_{\mathbf{T}}$, $\mathcal{F}_{\mathbf{AT}}$, $\mathcal{F}_{\mathbf{WT}}$) associated with the thigh-mounted IMU are also illustrated. The ultimate goal is to estimate the 3D rotations (FE, IE, AA) across the knee and doing so requires estimating the orientation of the shank anatomical frame, $\mathcal{F}_{\mathbf{AS}}$, relative to the thigh anatomical frame, $\mathcal{F}_{\mathbf{AT}}$. That critical step, however, requires introducing a *common* world frame of reference for the two IMUs as described below. Prior to that, the separate world frames and anatomical frames for both segments are established and described below.

The quaternion output (which is provided by proprietary software from APDM; for examples on how to compute quaternion output from IMU data, refer to [70, 71, 75, 76]) from an IMU, say $\mathbf{S}$, relates the orientation between an IMU frame, $\mathcal{F}_{\mathbf{S}}$, and a world frame, $\mathcal{F}_{\mathbf{WS}}$. The world frame $\mathcal{F}_{\mathbf{WS}}$ is defined by three mutually orthogonal axes ($\hat{\boldsymbol{X}}_{WS}$, $\hat{\boldsymbol{Y}}_{WS}$, $\hat{\boldsymbol{Z}}_{WS}$) with the $\hat{\boldsymbol{Z}}_{WS}$ axis chosen to align with gravity (using accelerometer data from $\mathbf{S}$), the $\hat{\boldsymbol{X}}_{WS}$ axis chosen to align with magnetic North (using magnetometer data from $\mathbf{S}$), and the $\hat{\boldsymbol{Y}}_{WS}$ axis computed from $\hat{\boldsymbol{Y}}_{WS} = \hat{\boldsymbol{Z}}_{WS} \times \hat{\boldsymbol{X}}_{WS}$ and thus chosen to point west. Let $\boldsymbol{R}_{WS/S}$ represent the resulting rotation matrix from $\mathcal{F}_{\mathbf{S}}$ to $\mathcal{F}_{\mathbf{WS}}$, a result that necessarily utilizes the magnetic North estimate from $\mathbf{S}$. An analogous procedure holds for the thigh-mounted IMU leading to the construction of $\boldsymbol{R}_{WT/T}$ representing the rotation matrix from $\mathcal{F}_{\mathbf{T}}$ to $\mathcal{F}_{\mathbf{WT}}$, a result that necessarily utilizes the magnetic North estimate from $\mathbf{T}$. Note also that the location and orientation of either IMU on its respective segment are arbitrary and the orientation of each IMU relative to its respective anatomical axes is established using functional alignment movements as detailed next.

The two functional alignment movements establish the shank anatomical axes ($\hat{\boldsymbol{X}}_{AS}$, $\hat{\boldsymbol{Y}}_{AS}$, $\hat{\boldsymbol{Z}}_{AS}$) that define $\mathcal{F}_{\mathbf{AS}}$ and the thigh anatomical axes ($\hat{\boldsymbol{X}}_{AT}$, $\hat{\boldsymbol{Y}}_{AT}$, $\hat{\boldsymbol{Z}}_{AT}$) that define $\mathcal{F}_{\mathbf{AT}}$. The procedure for both body segments is identical, and so only that

for the shank is detailed. First, the average acceleration measured during the still period yields a first estimate of the shank-fixed $\hat{\boldsymbol{Z}}_{AS}$ axis (superior-inferior axis) of the shank anatomical frame that is approximately aligned with gravity. (Note that during a trial with a human subject standing still, the direction of gravity remains approximately aligned with the superior-inferior axis.) This still period is followed by rotations purely about the CMM flexion-extension axis (while locking the other two rotational degrees of freedom). Following Seel et al. [19], the resulting angular velocity $\boldsymbol{\omega}_S$ and $\boldsymbol{\omega}_T$ measured by **S** and **T**, respectively, is used to define a unit vector aligned with the medial-lateral (hinge) axis measured in $\mathcal{F}_{\mathbf{S}}$. The medially pointing direction of this unit vector defines the anatomical axis. The anterior-posterior axis follows immediately from

$$\hat{\boldsymbol{Y}}_{AS} = \hat{\boldsymbol{Z}}_{AS} \times \hat{\boldsymbol{X}}_{AS} \tag{2.1}$$

Finally, the superior-inferior axis is adjusted (if needed) so that the medial-lateral axis remains orthogonal to the other two anatomical axes per,

$$\hat{\boldsymbol{Z}}_{AS} = \hat{\boldsymbol{X}}_{AS} \times \hat{\boldsymbol{Y}}_{AS} \tag{2.2}$$

The resulting orthonormal triad ($\hat{\boldsymbol{X}}_{AS}$, $\hat{\boldsymbol{Y}}_{AS}$, $\hat{\boldsymbol{Z}}_{AS}$), which are measured with respect to the shank IMU frame $\mathcal{F}_{\mathbf{S}}$, define the shank anatomical frame with the hinge axis $\hat{\boldsymbol{n}}_S = \hat{\boldsymbol{X}}_{AS}$. The (constant) rotation matrix from the shank IMU frame, $\mathcal{F}_{\mathbf{S}}$, to the shank anatomical frame, $\mathcal{F}_{\mathbf{AS}}$, follows from,

$$\boldsymbol{R}_{AS/S} = \begin{bmatrix} \hat{\boldsymbol{X}}_{AS} \\ \hat{\boldsymbol{Y}}_{AS} \\ \hat{\boldsymbol{Z}}_{AS} \end{bmatrix} \tag{2.3}$$

where each row contains the components of the anatomical axes measured with respect to $\mathcal{F}_{\mathbf{S}}$. Per ISB convention [18], the medial-lateral axis corresponds to the FE axis, the anterior-posterior axis corresponds to the AA axis, and the superior-inferior axis corresponds to the IE axis. An analogous procedure establishes the knee hinge axis $\hat{\boldsymbol{n}}_T = \hat{\boldsymbol{X}}_{AT}$ and the thigh anatomical frame ($\hat{\boldsymbol{X}}_{AT}$, $\hat{\boldsymbol{Y}}_{AT}$, $\hat{\boldsymbol{Z}}_{AT}$), which are

measured with respect to the thigh IMU frame $\mathcal{F}_{\mathbf{T}}$. The (constant) rotation matrix from the thigh IMU frame, $\mathcal{F}_{\mathbf{T}}$, to the thigh anatomical frame, $\mathcal{F}_{\mathbf{AT}}$, follows from,

$$\boldsymbol{R}_{AT/T} = \begin{bmatrix} \hat{\boldsymbol{X}}_{AT} \\ \hat{\boldsymbol{Y}}_{AT} \\ \hat{\boldsymbol{Z}}_{AT} \end{bmatrix} \tag{2.4}$$

2.3.3 Estimating 3D Knee Rotations Following Construction of a Common World Frame

The dynamic 3D knee rotations are ultimately estimated from the (time-varying) rotation matrix, $\boldsymbol{R}(t)_{AT/AS}$, from the shank anatomical frame, $\mathcal{F}_{\mathbf{AS}}$, to the thigh anatomical frame, $\mathcal{F}_{\mathbf{AT}}$. One may believe that this rotation matrix follows from the component rotations defined above per

$$\boldsymbol{R}(t)_{AT/AS} = \boldsymbol{R}_{AT/T}\boldsymbol{R}(t)_{T/WT}\boldsymbol{R}(t)_{WS/S}\boldsymbol{R}_{S/AS} \tag{2.5}$$

However, this result is correct only in the rare instances when the two IMU world frames, $\mathcal{F}_{\mathbf{WS}}$ and $\mathcal{F}_{\mathbf{WT}}$, are aligned. Frequently, the magnetometers in the two IMUs, **S** and **T**, provide distinct estimates of magnetic North (especially indoors where ferromagnetic interferences are often prevalent), and thus their respective world frames $\mathcal{F}_{\mathbf{WS}}$ and $\mathcal{F}_{\mathbf{WT}}$ will be misaligned in general. This is also true for methodologies that do not employ magnetometers, where changes in orientation are estimated from world frames constructed by other means/assumptions. Furthermore, the estimated world frames often vary with time due to sensor drift (bias) errors. In short, the two IMUs are independent sensors yielding independent and time-varying (drifting) world frames as also illustrated in the example results that follow.

Therefore, the key challenge here lies in constructing a common world frame of reference for the two IMUs, or equivalently, estimating the "correction" rotation matrix $\boldsymbol{C}(t)_{WT/WS}$ from $\mathcal{F}_{\mathbf{WT}}$ to $\mathcal{F}_{\mathbf{WS}}$. There is no clear way of determining the angular differences in the estimates of magnetic North from the two IMUs. However, one can exploit the constraint that the hinge (medial-lateral) axes, $\hat{\boldsymbol{n}}_S$ and $\hat{\boldsymbol{n}}_T$, should

be identical in a common world frame during the time intervals when the knee can reliably simplified to a hinge joint. This "hinge constraint" is treated as a full vector equation and therefore generalize the scalar treatment of this constraint employed in [19]. Whenever this hinge criterion is satisfied, the correction rotation matrix $\boldsymbol{C}(t)_{WT/WS}$ can be constructed from $\hat{\boldsymbol{n}}_S$ and $\hat{\boldsymbol{n}}_T$ by computing the axis of rotation, $\hat{\boldsymbol{k}}$, and the associated angle of rotation, θ, needed to align and in their respective world frames. To this end, the cross product,

$$\hat{\boldsymbol{k}}(t) = \boldsymbol{R}(t)_{WS/S}\hat{\boldsymbol{n}}_S \times \boldsymbol{R}(t)_{WT/T}\hat{\boldsymbol{n}}_T \tag{2.6}$$

defines the axis of rotation and the dot product,

$$\theta(t) = \cos^{-1}\left(\boldsymbol{R}(t)_{WS/S}\hat{\boldsymbol{n}}_S \cdot \boldsymbol{R}(t)_{WT/T}\hat{\boldsymbol{n}}_T\right) \tag{2.7}$$

defines the required angle of rotation. The requisite correction rotation matrix follows from Rodrigues rotation formula [83] per

$$\boldsymbol{C}(t)_{WT/WS} = \boldsymbol{I} + \sin\theta(t)\hat{\boldsymbol{k}}^{\times}(t) + \big(1 - \cos\theta(t)\big)\hat{\boldsymbol{k}}^{\times}(t)^2 \tag{2.8}$$

where $\boldsymbol{I}$ is the identity matrix and $\hat{\boldsymbol{k}}^{\times}$ is the skew symmetric form of $\hat{\boldsymbol{k}}$. The correction rotation matrix corrects the small misalignment between the two world frames and between successive times when the hinge criterion (described below) is satisfied. The resulting corrected form of Eqn. 2.5 becomes

$$\boldsymbol{R}(t)_{AT/AS} = \boldsymbol{R}_{AT/T}\boldsymbol{R}(t)_{T/WT}\boldsymbol{C}(t)_{WT/WS}\boldsymbol{R}(t)_{WS/S}\boldsymbol{R}_{S/AS} \tag{2.9}$$

where $\boldsymbol{R}(t)_{AT/AS}$ is again the needed time-varying rotation matrix describing the orientation of the anatomical shank frame relative to the anatomical thigh frame. Note that the ISB recommends [18] first calculating FE, then IE, and finally AA. However, due to the mechanical design of the knee analog, the reverse order is required. For the knee analog, the joints are in series order such that the rotation sequence needed to rotate from the shank link to the thigh link require rotations about the AA axis

first, the IE axis second, and the FE axis third. Decomposing with that order in mind (analogous to the procedure in [75]) yields the 3D rotation angles across the knee analog.

The above strategy for aligning the two world frames holds for the time intervals when the knee analog in this experiment is predominantly functioning as a hinge joint. To this end, one must develop criteria for determining: 1) when the knee is predominantly functioning as a hinge joint and, 2) a strategy for aligning the two world frames in between those time intervals (i.e., when the knee is no longer predominantly functioning as a hinge joint).

The time intervals when the knee is predominantly functioning as a hinge joint are identified under two conditions: Case 1a) when the knee joint is approximately stationary, and Case 1b) when the knee joint is rotating.

1a) Functioning as a Hinge - Stationary Case: The knee analog may be considered a hinge joint in the limiting case when it is essentially locked (i.e. straight) and stationary. These time intervals are identified when the segments are stationary and also nominally aligned with gravity (such as the still period in the functional alignment movement sequence). In such instances, the knee functions as a hinge with zero hinge rotation rate. These conditions are considered satisfied in these experiments when

$$\max\left\{ \left\|\boldsymbol{a}_S - \boldsymbol{g}\right\|, \left\|\boldsymbol{a}_T - \boldsymbol{g}\right\| \right\} \leq 0.02 \left\|\boldsymbol{g}\right\| \tag{2.10}$$

$$\text{mean}\left\{ \cos^{-1}\left(\frac{\boldsymbol{a}(t)_S \cdot \boldsymbol{a}(0)_S}{\left\|\boldsymbol{a}(t)_S\right\| \left\|\boldsymbol{a}(0)_S\right\|} \right), \cos^{-1}\left(\frac{\boldsymbol{a}(t)_T \cdot \boldsymbol{a}(0)_T}{\left\|\boldsymbol{a}(t)_T\right\| \left\|\boldsymbol{a}(0)_T\right\|} \right) \right\} \leq 3^\circ \tag{2.11}$$

in which $\boldsymbol{a}(0)_S$ and $\boldsymbol{a}(0)_T$ denote the (constant) acceleration measured during the functional alignment step with the segments vertical.

1b) Functioning as a Hinge - Rotating Case: When the conditions for the stationary case above are not met, the knee joint is considered rotating. During such instances, the knee may still function predominantly as a hinge joint (with non-zero hinge rotation rate) whenever the angular velocities of the shank and

the thigh are predominantly aligned with the hinge axis defined by and on each segment respectively. For this experiment, the knee functions as a hinge with non-zero hinge rotation rate whenever

$$\min\left\{|\boldsymbol{\omega}_S|, |\boldsymbol{\omega}_T|\right\} > 30^\circ/s \tag{2.12}$$

$$\text{mean}\left\{\frac{\boldsymbol{\omega}_S \cdot \hat{\boldsymbol{n}}_S}{|\boldsymbol{\omega}_S|}, \frac{\boldsymbol{\omega}_T \cdot \hat{\boldsymbol{n}}_T}{|\boldsymbol{\omega}_T|}\right\} \geq 0.99 \tag{2.13}$$

The numerical thresholds for the criteria above are stringent, but appropriate for the knee analog employed in the study. There are frequent periods of times when neither criteria are met and the knee analog no longer functions purely as a hinge as described next.

2) Not Functioning as a Hinge: When neither case above holds, the knee is no longer functioning as a hinge and appreciable internal-external rotation and/or abduction-adduction exists. During such time intervals, it is assumed that varies slowly and continuously and that it can be estimated by linear interpolation between two consecutive "update" times when either of the two cases above hold.

2.4 Results and Discussion

2.4.1 Qualitative Comparisons

The rotation matrices between the sensor frames and their respective world frames are defined by both horizontal (or yaw) and vertical (or elevation) angular components. Studying these angular components is important for understanding the challenge (and the solution) to constructing the correction rotation matrix $\boldsymbol{C}_{WT/WS}$ introduced above that corrects the differing and drifting IMU world frames.

Figure 2.2 illustrates the orientation of the shank IMU frame $\mathcal{F}_{\mathbf{S}}$ relative to its world frame $\mathcal{F}_{\mathbf{WS}}$ in terms of both yaw ψ and vertical ϕ angular components. In particular, the unit vectors $\hat{\boldsymbol{X}}_{WS}$ and $\hat{\boldsymbol{Y}}_{WS}$ define the horizontal plane for $\mathcal{F}_{\mathbf{WS}}$, which

differ from that of $\mathcal{F}_\mathbf{S}$ by the vertical angle ϕ. Moreover, the frames also differ by the yaw angle ψ, which is defined as the angle between $\hat{\boldsymbol{X}}_{WS}$ and the projection of $\hat{\boldsymbol{x}}_S$ onto the horizontal plane for $\mathcal{F}_\mathbf{WS}$. One way to visualize the drift error is to examine how either angle varies with time over a trial during which the knee is periodically returned to its initial orientation.

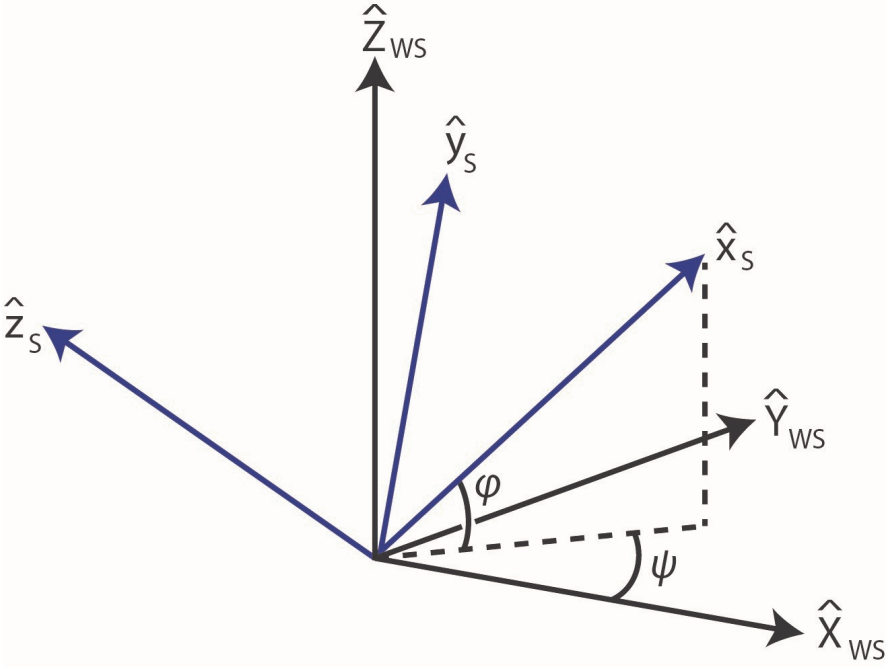


Figure 2.2: The orientation of the shank IMU frame $\mathcal{F}_\mathbf{S}$ (blue) relative to the shank world frame $\mathcal{F}_\mathbf{WS}$ (black). The orientation is defined by the illustrated yaw ψ and vertical ϕ angular components. The horizontal dotted line is the projection of $\hat{\boldsymbol{x}}_S$ onto the $\hat{\boldsymbol{X}}_S$-$\hat{\boldsymbol{Y}}_S$ plane and the vertical dotted line is the projection onto the vertical direction.

For example, Figure 2.3a illustrates how the yaw angles computed for both IMUs (relative to their respective world frames) vary with time for the simplest testing session when the knee undergoes pure flexion-extension. The portion of the testing session shown encompasses the two functional alignment movements (first shaded interval corresponds to the still period and the first unshaded interval corresponds to FE rotations), followed by another still period (second shaded interval), then

a longer (second unshaded) interval with four trials of five repetitive knee flexion-extension movements (with 46 more trials thereafter that are not shown). The knee is approximately returned to its original position at the end of each knee flexion-extension trial. Note that this approximately 3-minute sample is taken from a testing session lasting approximately 15 minutes, and substantial drift arises over this long period of time. Figure 2.3b illustrates the Boolean values for the criteria for the two cases (Case 1a and Case 1b) when the knee acts as a hinge; a value of 1 corresponding to the criteria being satisfied and a value of 0 corresponding to the criteria not being satisfied. As expected, the knee analog functions as a hinge for the stationary case (Case 1a) during the rest periods. Similarly, the knee analog functions as a hinge during the subsequent rotations that induce pure flexion-extension for this trial (Case 1b). (There are also short intervals when Case 1b is not satisfied as the angular velocity magnitudes fall below the selected threshold values.)

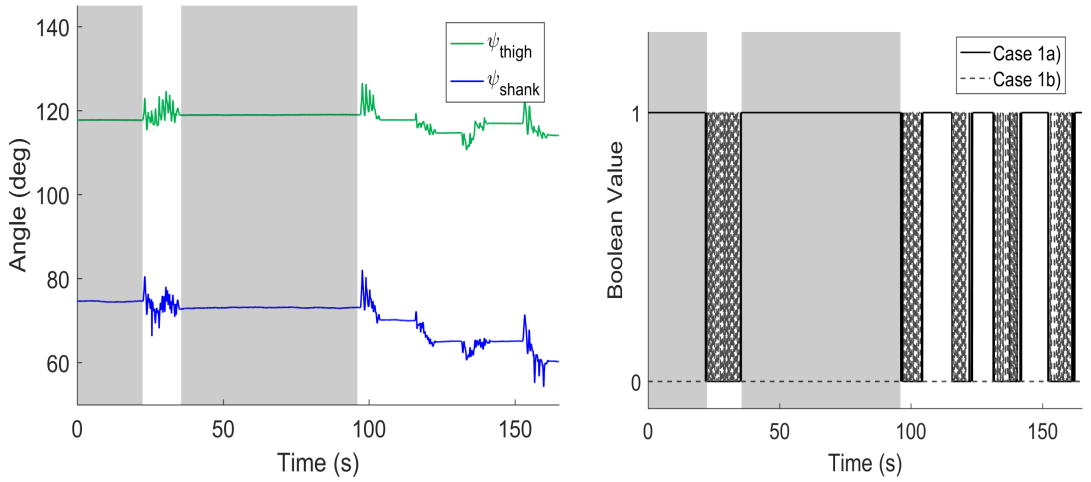


Figure 2.3: Example results from pure flexion-extension trial. **(a)** Yaw angles for shank- and thigh-mounted IMU versus time encompassing two functional alignment movements (first shaded and unshaded regions, respectively), a rest period (second shaded region), and then four trials of five repetitive knee flexion-extension movements between which the knee is returned to the original position. **(b)** Boolean (0 or 1) values for criteria defining Case 1a (stationary) and Case 1b (rotating) for which the knee analog acts as a hinge. The solid black line is for Case 1a and the dashed grey line is for Case 1b. The shaded and unshaded areas denote the same regions in **(a)**.

The orientation drift error manifests in the slowly changing (low frequency) components of ψ_{thigh} and ψ_{shank} despite the fact that the knee returns to its nominal orientation between successive flexion-extension movements. In particular, the net change in ψ_{thigh} and ψ_{shank} over the entire 15-minute period are 65° and 101°, respectively. This corresponds to drift rates for the thigh and shank sensors of -0.07°/s and -0.11°/s, respectively, which are consistent with previously measured drift rates for MEMS inertial sensors; see, for example, [84]. The two world frames are misaligned at the start of the trial and continue to drift apart throughout the trial and it should also be noted the drift is not exclusively about the vertical axis. The misalignment of the world frames, and its associated drift, is compounded by sources of ferromagnetic interference in the laboratory environment. Despite the observably

large misalignment of the world frames, the method above correctly identifies the correction rotation matrix needed to accurately estimate 3D rotations across the knee analog.

Figure 2.4 illustrates the differences between the true (CMM) and estimated (IMU) flexion-extension, internal-external rotation, and abduction-adduction angles for one trial of five repetitive flexion-extension movements previously considered in Figure 2.3. These estimates closely track the truth values reported by the optical encoders; refer also to quantitative comparison that follows. Also shown are the differences between the true and (uncorrected) estimates that result from using Eqn. 2.5 instead of Eqn. 2.9; or equivalently, assuming that $C_{WT/WS} = I$. Clearly, ignoring this correction leads to large errors and for all three rotation angles. Note that throughout this trial, the knee analog functions largely as a hinge, meaning that either Case 1a) or Case 1b) is almost always satisfied (and the linear interpolation associated with Case 2 is not required); refer to results illustrated in Figure 2.3b. However, this is not the case in the following trials that induce substantial internal-external rotation and/or abduction-adduction.

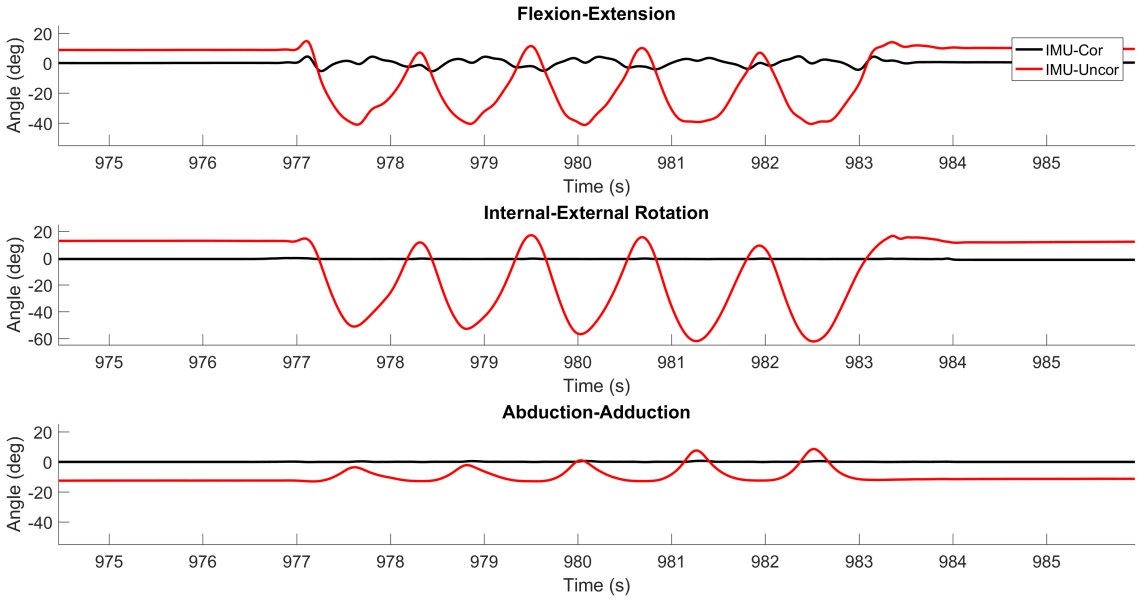
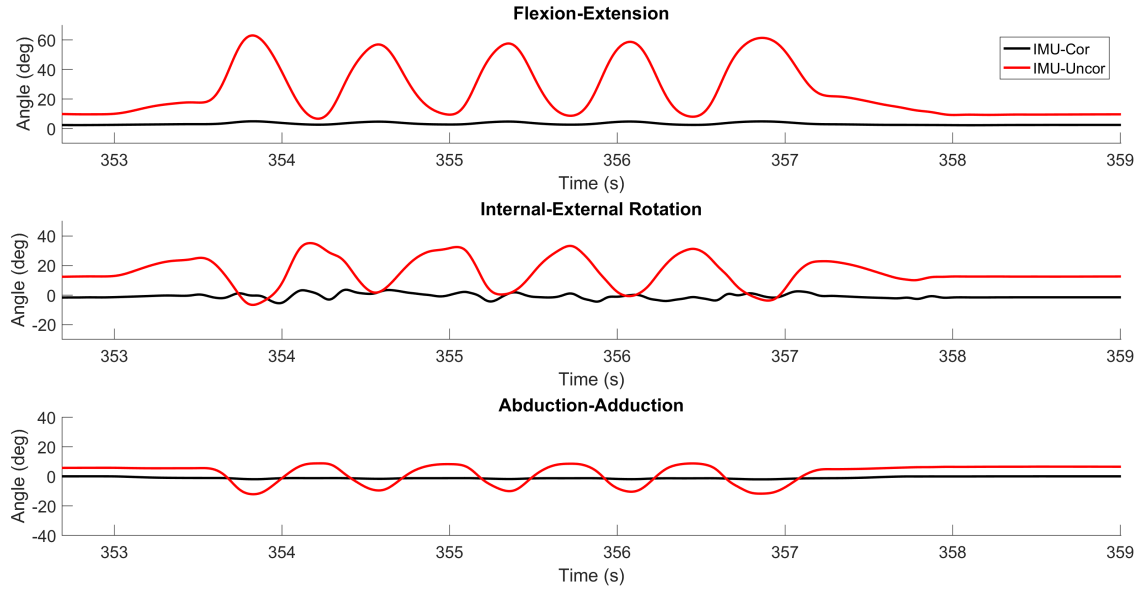
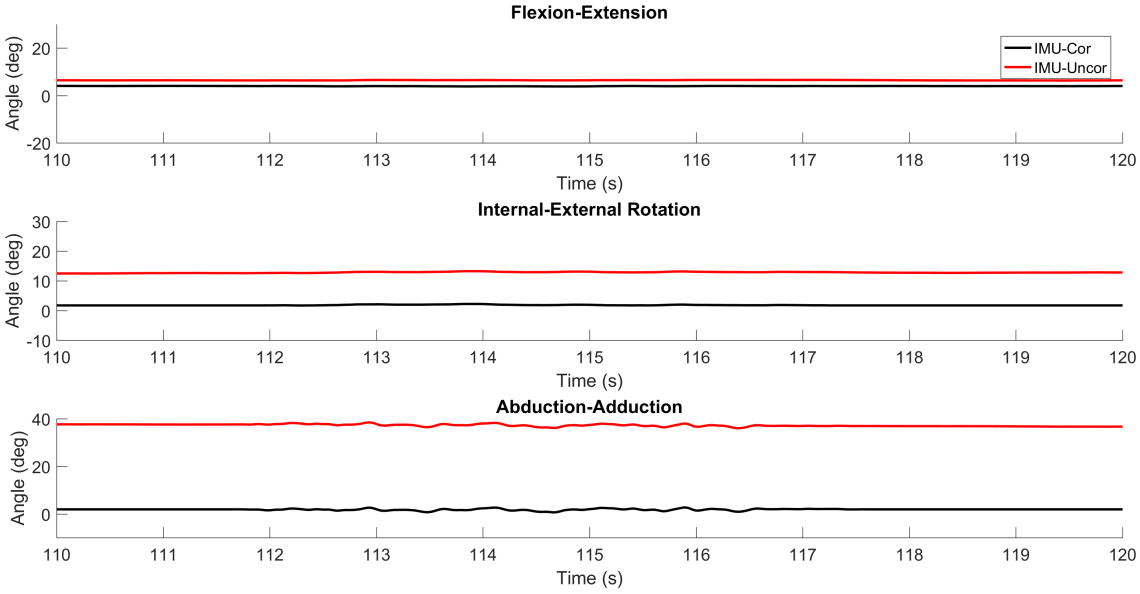


Figure 2.4: Example results from pure flexion-extension trial. The difference between CMM truth data and IMU-derived estimates with the correction (black) and without the correction (red) of flexion-extension, internal-external rotation, and abduction-adduction are plotted for a representative time period.

Figure 2.5 illustrates example results for two trials that consider (a) pure internal-external rotation and (b) pure abduction-adduction. As with the prior case, the differences between the true and estimated flexion-extension, internal-external rotation, and abduction-adduction angles remain small, meaning the estimates closely track the truth values reported by the optical encoders. For instance, observe the very slight off-axis errors along rotation axes that are otherwise physically constrained on the knee analog. These small off-axis rotations likely derive from very minor errors in the estimated orientations of one or both of the anatomical frames, $\mathcal{F}_{\mathbf{AS}}$ and $\mathcal{F}_{\mathbf{AT}}$. Nonetheless, the agreement between the IMU estimates and the truth data remains excellent. By contrast, large errors again arise in general when one ignores the correction. The agreement with the correction for these limiting cases of 1D rotation remains undiminished when the knee analog is unconstrained and undergoes fully 3D rotations as shown next.



(a)



(b)

Figure 2.5: Example results from **(a)** pure internal-external rotation trial, and **(b)** pure abduction-adduction trial. The differences between the CMM truth data and IMU-derived estimates with the correction (black) and without the correction (red) of flexion-extension, internal-external rotation, and abduction-adduction are plotted for a representative time period.

First, a sample time record of the yaw angles for each IMU is presented in Figure 2.6a when the knee undergoes combined 3D rotations. Following Figure 2.3, Figure 2.6 includes a portion of a testing session that encompasses the two functional alignment movements (first shaded and unshaded intervals, respectively), followed by a nominally still period during which the constraints are removed (second shaded interval), then a longer interval with six repetitive 3D movements (with four more thereafter that are not shown). The knee analog is approximately returned to its original position at the end of each knee movement. Note that this is taken from a testing session lasting approximately 6 minutes, and substantial drift arises over this period of time. Compared to the results of Figure 2.3, the results of this longer time sample in Figure 2.6a exhibit even greater drift error as again manifested in the slowly changing (low frequency) components of ψ_{thigh} and ψ_{shank}. The net changes in ψ_{thigh} and ψ_{shank} over the entire 6-minute period is 46° and 28°, respectively, which correspond to drift rates of 0.14°/s and -0.08°/s. Despite the very apparent drift, the method produces excellent estimates of the 3D rotations. Figure 2.6b documents the Boolean values for the criteria for the two cases (Case 1a and Case 1b) when the knee acts as a hinge; a value of 1 corresponding to the criteria being satisfied and a value of 0 corresponding to the criteria not being satisfied. As before, the knee analog functions as a hinge for the stationary case (Case 1a) during rest periods and functions as a hinge during the flexion-extension functional alignment movement (Case 1b). Note that during the second shaded area, the physical constraints are being removed from the CMM in preparation for the combination rotation trials to follow.

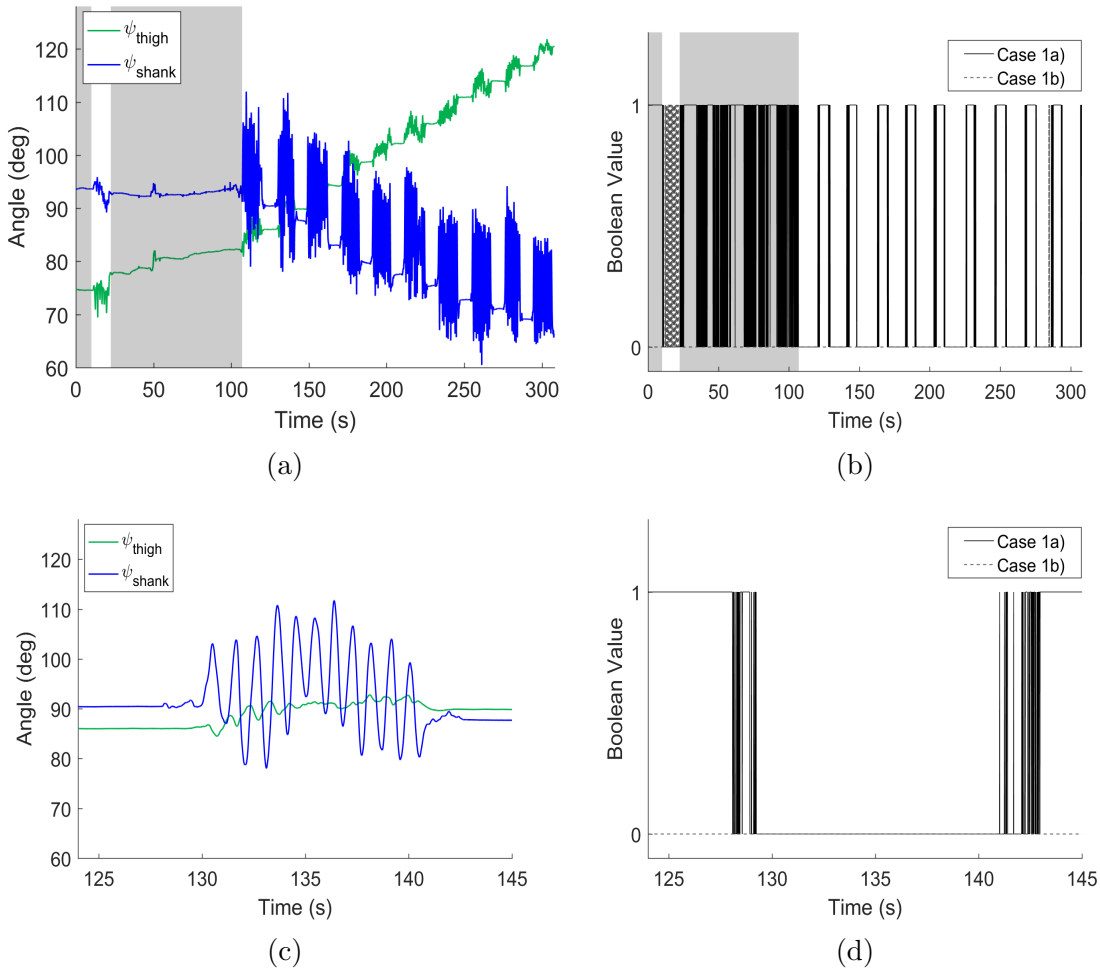


Figure 2.6: Example results from combined 3D rotation trial. **(a)** Yaw angles for shank- and thigh-mounted IMU versus time encompassing two functional alignment movements (first shaded and unshaded regions, respectively), a nominal rest period during which the constraints are removed (second shaded region), and then ten repetitive 3D movements between which the knee is approximately returned to the original position. **(b)** Boolean (0 or 1) values for criteria defining Case 1a (stationary) and Case 1b (rotating) for which the knee analog acts as a hinge. The solid black line is for Case 1a and the dashed grey line is for Case 1b. The shaded and unshaded areas denote the same regions in **(a)**. In addition, **(c)** and **(d)** illustrate sample results from **(a)** and **(b)**, respectively, on a fine (second-level) time scale.

Figure 2.7 illustrates the estimated flexion-extension, internal-external rotation, and abduction-adduction angles for a representative time period from Figure 2.6 during which all three angles were being simultaneously exercised. Consistent with the above results, the estimates of all three angles closely track the truth values reported by the optical encoders when the correction is employed. When the correction is not employed, the estimates can be rather poor and particularly so for abduction-adduction in this example.

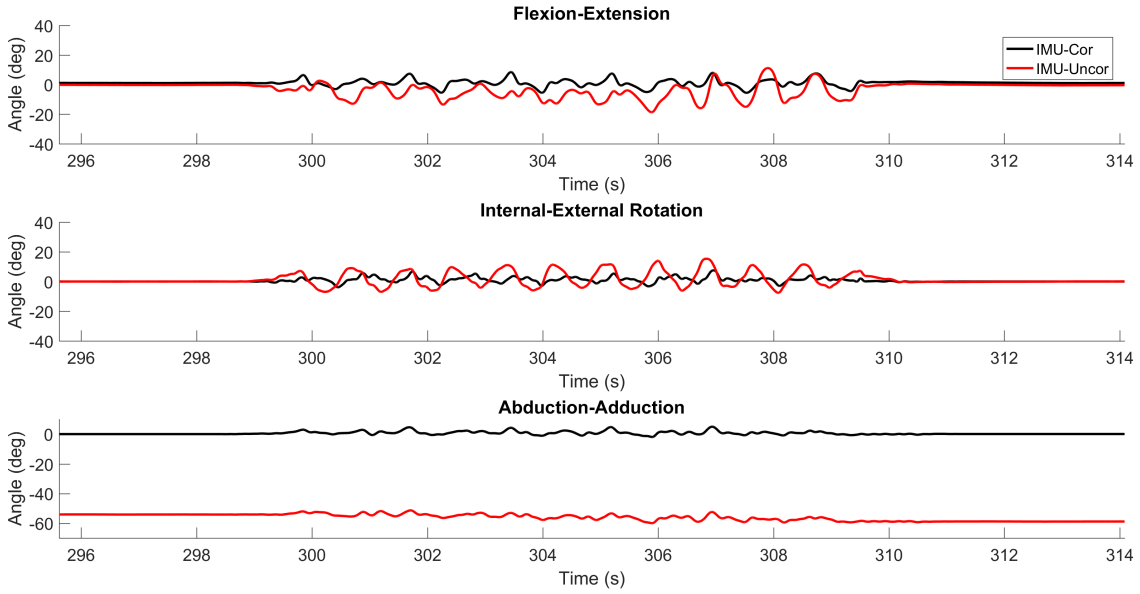


Figure 2.7: Example results from combined 3D rotation trial. The differences between the CMM truth data and the IMU-derived estimates with the correction (black) and without the correction (red) for flexion-extension, internal-external rotation, and abduction-adduction are plotted for a representative time period.

2.4.2 Quantitative Comparisons

The above results illustrate very close qualitative agreement between the IMU-derived estimates of the joint angles and those measured directly using the embedded optical encoders. Next, quantitative comparisons for the entire data set are provided and discussed.

To start, consider Figure 2.8 which shows the IMU-estimated flexion-extension angle versus that measured by the optical encoder for the duration of all 5 testing sessions (cumulatively about 30 minutes) of combined 3D rotations. The best fit line to this data yields a slope of 1.01, a y-intercept of $b = 0.07°$, and a correlation coefficient of $r = 0.99$. Thus, the estimates exhibit extremely high correlation with the truth data (and just slightly over predicting flexion-extension relative to the truth data). In addition, the root-mean square (RMS) error between the estimates and the truth data is 3.46° or 2.96% relative to the 117° range of motion. These results, and the analogous quantitative comparisons for all of the experiments, are summarized in Tables 2.1 and 2.2.

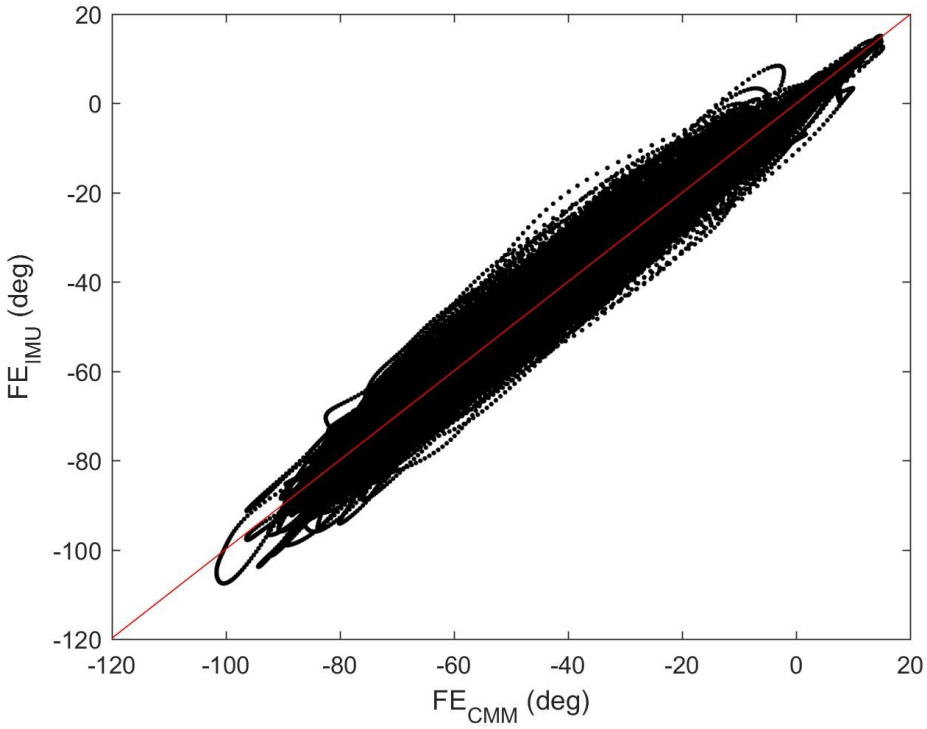


Figure 2.8: Correlation of the results for the combined 3D rotation trials. IMU-estimated flexion-extension plotted against the corresponding truth data from optical encoder. The red line is the linear fit.

Table 2.1: Quantitative comparisons for cases of pure rotation about a single axis including range of motion (ROM), RMS error, correlation (r), and slope, and y-intercept (b) of linear fit.

1D Rotation	**ROM** (°)	**RMS Error** (°)	$\boldsymbol{r}$	**Slope**	$\boldsymbol{b}$ (°)
Pure Flexion-Extension	161	3.90	0.99	1.01	0.12
Pure Internal-External	72.8	1.83	0.99	1.01	0.04
Pure Abduction-Adduction	17.0	0.12	0.99	0.99	-0.02

Table 2.2: Quantitative comparisons for the combined 3D rotation trial including range of motion (ROM), RMS error, correlation (r), slope, and y-intercept (b) of linear fit.

3D Rotation	**ROM** (°)	**RMS Error** (°)	$\boldsymbol{r}$	**Slope**	$\boldsymbol{b}$ (°)
Flexion-Extension	117	3.46	0.99	1.00	0.07
Internal-External	98.4	2.48	0.99	1.02	0.06
Abduction-Adduction	58.3	1.69	0.94	0.92	0.04

Table 2.1 reports the quantitative comparisons of IMU-estimated angles to those measured by the optical encoders for the three limiting cases of pure rotation about the FE, IE and AA axes. Reported are the range of motion (ROM) about each axis, the RMS error between the estimated and measured angles, and the correlation coefficient, slope and y-intercept of the associated linear fit. Analogous results are reported in Table 2.2 for the combined 3D rotation movements. Regardless of the trial (pure rotation about any one axis or combined rotation about all three axes), the IMU-derived angle estimates remain within 4° of those measured by the optical encoders, have correlation coefficients exceeding 0.94 and slopes between 0.99 and 1.02. In short, the method yields estimates that replicate the truth data with high confidence. The ranges of motion for each angle are motivated by the behavior of the human knee for which flexion-extension typically has the greatest range of motion, followed by internal-external rotation, and abduction-adduction. However, the knee

analog is exercised over far greater ranges of motion than observable on a healthy human knee for the purpose of this thorough validation study.

2.5 Summary and Conclusions

This study contributes a new method to estimate the 3D rotations across the knee using a pair of shank- and thigh-mounted IMUs that yields highly accurate results when benchmarked on a coordinate measurement machine (CMM). Central to this method is constructing a common world frame of reference for the two IMUs by exploiting a vector constraint equation for the medial-lateral axis of the knee. This constraint arises during the time periods when the knee behaves predominantly as a hinge joint, during which and between times the common world frame can be constructed despite sensor orientation drift errors. It therefore has considerable potential for accurately estimating the 3D rotations across the human knee. Doing so will also require further research to address additional considerations. First, the sensor orientation estimates used in this work are provided by proprietary software. To more accurately study the efficacy of implementing the anatomical kinematic constraint as a world frame correction requires estimates of orientation, which is addressed in Chapter 3.

Next, unlike the CMM, the knee joint has some laxity which may also influence the estimated 3D joint angles. The stringent limits used in Eqns. 2.10-2.13, while fully appropriate for the CMM, must necessarily be relaxed for the human knee. Validations using human subjects will guide the selection of these new limits. Also, while the IMUs were rigidly fastened to the links of the CMM, the IMUs attached to the human shank and thigh may move slightly relative to the underlying skeleton due to soft tissue movements, a problem common to all motion capture methods. The anatomical significance of the estimated 3D rotations also depends on accurately establishing the orientation of the shank and thigh anatomical frames relative to their respective IMU frames as input to Eqn. 2.9. This motivates the need to study a variety of functional alignment movements to achieve this intermediate result, which will be discussed in Chapter 4.

CHAPTER 3

Measurement Theory for Estimating Knee Rotations with Probabilistic World Frame Alignment

3.1 Introduction

The objective of this chapter is to expand and further validate the measurement theory described in Chapter 2 via two important extensions. The first replaces the sensor orientation estimates previously obtained using proprietary software with estimates obtained by a customized Kalman filter. The second replaces the deterministic treatment of the anatomical kinematic constraint with a probabilistic treatment employing a custom Error-State Kalman Filter. The resulting method, based on (probabilistic) estimation theory, is superior in its ability to estimate three dimensional knee rotations during extended periods of dynamic movements like what is studied in Chapter 4.

Historical approaches in estimating orientation of IMUs fall into one of two categories, namely: 1) deterministic (complementary) or 2) probabilistic (stochastic). Arguably the most well-known deterministic approach is Madgwick's filter [85], which is a nonlinear observer that provides a computationally efficient estimate of orientation by solving a gradient descent optimization problem (and with a single iteration). However, the accuracy of this method requires that the convergence rate of the op-

timization being equal or greater than the rate of orientation change (i.e., angular velocity magnitude). As a result, the (fixed-step size) gradient descent optimization tends to lag the true state during sustained highly dynamic movements resulting in a tradeoff between computational cost and accuracy [86].

On the other hand, traditional probabilistic approaches, namely various types of indirect Kalman filtering, have the reciprocal strengths and weaknesses. These methods tend to be more computationally expensive due to their complexity [87], but provide accurate estimates even during prolonged dynamic movements provided the filter is parameterized properly (e.g., process and measurement noises are reasonably well-estimated) [86]. Furthermore, these formulations allow for integration of additional information to provide even more accurate orientation estimates. However, inappropriately updating the state estimation even once can cause the filter to diverge [88]. Therefore, external disturbances must be carefully considered to render the filter robust. For example, when the IMU is static (i.e., stationary), the accelerometer provides an estimated direction of gravity, which is used to correct the orientation relative to the horizontal plane (i.e., pitch and roll). Therefore, any acceleration of the rigid body to which the IMU is attached is superimposed on the acceleration due to gravity thereby polluting the estimated orientation. Similarly, an estimate of magnetic North is frequently integrated into the estimation procedure using magnetometer data, which is then used to correct orientation about vertical (i.e., yaw) [76]. However, magnetometer data is often polluted by magnetic interference from surrounding ferromagnetic materials in the local environment (e.g., equipment, building materials, piping, wiring, etc.) thereby yielding faulty estimates of magnetic North [89]. Many researchers have proposed distinct approaches to minimize the errors introduced by the accelerometer and/or magnetometer data (see, for example, [77, 90]).

The Kalman filter is *the* optimal estimator that minimizes the mean squared error when the state and measurement dynamics are modeled as linear and the process/measurement noise processes are modeled as white Gaussian [91]. However, the equations governing orientation estimates (like many other dynamical quantities) are nonlinear, which poses a challenge particularly for propagating the uncertainty over

time. Fortunately, several successful solutions to this challenge have been discovered. The first (and most common) is to use an Extended Kalman filter in which the nonlinear dynamical equations are linearized about the current estimated state via a Taylor series expansion using the Jacobian of the nonlinear dynamic (state transition) equations [90]. Another approach is to use a Sigma-Point (Unscented) Kalman filter in which a set of samples (sigma points) are taken from the distribution describing the current state, propagated to the current time step via the nonlinear dynamic (state transition) equations, and then averaged to provide an estimate of the current state [92]. These two approaches produce similar results with the Sigma-Point filter providing slightly better estimates, but it is computationally more expensive and tends to lag the true state more than the Extended Kalman filter [93]. Other noteworthy approaches also include the Particle filter [94], the Invariant Extended Kalman filter [95], and the Error-State Kalman filter (ESKF) [96], the last of which is pursued in this work.

The key idea underlying the error-state formulation of the Kalman filter is to decompose the true state of the system (i.e., the quaternion representation of the IMU's orientation in this case) into a nominal-state and an error-state [96]. The nominal-state arises from integrating the (potentially highly dynamic) IMU data without consideration of noise (uncertainty) or model imperfections. Thus, any errors resulting from that integration are attributed to the error-state, whose dynamic equations are *linear* with respect to the error-state. Specifically concerning IMU orientation, the benefits of the ESKF formulation are outlined by Madyastha et al. [91] and Solà [97] and briefly summarized here. The orientation error-state is a minimal parameterization because the scalar term of the quaternion (q_0) is assumed to be unity since the error in orientation is assumed small (i.e., $cos(\theta) \approx 1$ when θ is small). This approximation avoids singularities in the covariance matrices arising from enforced constraints (i.e., forcing the quaternions to have unit magnitudes). Next, the error-state is assumed small such that the linearization remains accurate. Finally, the error state dynamics are relatively slow compared to those of the (deterministic) nominal-state, which means corrections to the error-state can be made less frequently than in an EKF. Consequently, the criteria for integrating measurements

into the state estimation can be stricter to avoid divergence.

The remainder of this chapter reports two studies. The first provides a derivation of the ESKF formulation developed for a single IMU which is then validated with truth data provided by the same coordinate measurement machine (CMM) used in Chapter 2. The second extends the ESKF to two IMUs by simultaneously treating anatomical constraint of Chapter 2 as a probabilistic correction to the IMU world frame estimates. This algorithm is similarly validated using truth data from the CMM.

3.2 Study 1: Robust ESKF Validation for a Single IMU

3.2.1 Background on Kalman Filtering

This section provides a simplified description of Kalman filtering in the context of estimating IMU orientation, which draws largely from [98] and [91]. For more complete discussions of various types of Kalman filtering, see the descriptions provided by Welch and Bishop [99] or Daum [100].

Assume a state vector, $\tilde{\boldsymbol{x}}$, is a collection of stochastic (random) variables, say

$$\tilde{\boldsymbol{x}} = \begin{bmatrix} \tilde{\boldsymbol{x}}_1 \\ \tilde{\boldsymbol{x}}_2 \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{q}} \\ \tilde{\boldsymbol{\omega}}_b \end{bmatrix} \tag{3.1}$$

where $\tilde{\boldsymbol{q}}$ is a quaternion and $\tilde{\boldsymbol{\omega}}_b$ are the angular rate gyro biases. The probability distribution of $\tilde{\boldsymbol{x}}$, $P(\tilde{\boldsymbol{x}})$, maps probabilities to all possible realizations of $\tilde{\boldsymbol{x}}$. The expectation of $\tilde{\boldsymbol{x}}$ (denoted $\varepsilon[\tilde{\boldsymbol{x}}]$) is the expected value of the state vector, which represents the mean of $\tilde{\boldsymbol{x}}$ ($\boldsymbol{\mu}_x$) otherwise known as the 1^{st} moment of $\tilde{\boldsymbol{x}}$. A moment is a quantitative measure of the shape of a function, and probability distributions are frequently described in terms of their moments (1^{st}: mean=$\boldsymbol{\mu}_x = \varepsilon[\tilde{\boldsymbol{x}}]$, 2^{nd}: variance=$\varepsilon[(\boldsymbol{\mu}_x - \tilde{\boldsymbol{x}})^2]$, 3^{rd}: skewness=$\varepsilon[(\boldsymbol{\mu}_x - \tilde{\boldsymbol{x}})^3]$, 4^{th}: kurtosis=$\varepsilon[(\boldsymbol{\mu}_x - \tilde{\boldsymbol{x}})^4]$, ...). Normally distributed (Gaussian) random variables are special in that they can be fully characterized by their first two moments, namely the mean and variance.

As mentioned above, the Kalman filter yields the minimum mean squared error

filter in the following sense; for a more thorough discussion, see [101]. Given the true expected value is $\varepsilon[\tilde{\boldsymbol{x}}] = \boldsymbol{\mu}_x$ and the estimated expected value is $\varepsilon[\hat{\tilde{\boldsymbol{x}}}] = \hat{\boldsymbol{\mu}}_x$, the expected error $\boldsymbol{e}$ is defined as the innovation of the true and estimated expected values per $\boldsymbol{e} = \boldsymbol{\mu}_x - \hat{\boldsymbol{\mu}}_x$. The best guess in describing $P(\tilde{\boldsymbol{x}})$ is to assume $\tilde{\boldsymbol{e}}$ is normally distributed about the estimated expected value, $\hat{\boldsymbol{\mu}}_x$ (hereafter referred to as $\hat{\boldsymbol{x}}$ for simplicity), with an (error) covariance matrix, $\boldsymbol{P}$, describing the uncertainty associated with that estimated expected value. Since the true state is assumed to be an unobserved Markov process (i.e., the current state is conditioned on the previous state only), Kalman filters easily implemented as recursive algorithms that consist of two steps, namely: 1) predict (propagate) and 2) update (correct).

Step 1: Predict

This step employs solely the data from the angular rate gyro to propagate the estimated state from the previous time step to that of the current time step. This step starts with the process model governing the dynamics of the system per

$$\tilde{\boldsymbol{x}}_k = f(\tilde{\boldsymbol{x}}_{k-1}, \boldsymbol{u}_k) + \tilde{\boldsymbol{w}}_k \tag{3.2}$$

where k is the current (discrete) time step, $\boldsymbol{u}$ is the (deterministic) input to the process model, and $\tilde{\boldsymbol{w}}$ is the unmodeled system uncertainty (process noise) which is assumed Gaussian, uncorrelated, and with zero mean. Assuming the process described by $f(\cdot)$ is linear, taking the expectation yields

$$\hat{\boldsymbol{x}}_k = \boldsymbol{A}\hat{\boldsymbol{x}}_{k-1} + \boldsymbol{B}\boldsymbol{u}_k \tag{3.3}$$

where $\boldsymbol{A}$ represents the (time-invariant) state transition matrix relating the previous state to the current state and $\boldsymbol{B}$ represents the (time-invariant) control input matrix relating the input to the current state. A similarity transform propagates the estimated error covariance matrix, $\hat{\boldsymbol{P}}$, from the previous time step to the current time step per

$$\hat{\boldsymbol{P}}_k = \boldsymbol{A}\hat{\boldsymbol{P}}_{k-1}\boldsymbol{A}^T + \boldsymbol{Q} \tag{3.4}$$

where $\boldsymbol{Q}$ is the process noise matrix associated with $\tilde{\boldsymbol{w}}$ from Eqn. 3.2.

However, the estimated states are a quaternion, $\boldsymbol{q}$, describing the (potentially rapidly changing) orientation of the IMU and the angular rate gryo biases, $\boldsymbol{\omega}_b$, representing the slowly varying dc offsets in the angular velocity measurements. The differential equation governing the dynamics of the system are

$$\dot{\hat{\boldsymbol{x}}}(t) = \begin{bmatrix} \dot{\hat{\boldsymbol{q}}}(t) \\ \dot{\hat{\boldsymbol{\omega}}}_b(t) \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\boldsymbol{q}(t) \otimes (\boldsymbol{\omega}(t) - \boldsymbol{\omega}_b(t) - \tilde{\boldsymbol{\omega}}_n) \\ \tilde{\boldsymbol{\omega}}_w \end{bmatrix} \tag{3.5}$$

are based solely on the angular rate gyro data $\boldsymbol{\omega}(t)$ where $\otimes$ represents the quaternion product, $\tilde{\boldsymbol{\omega}}_n$ is the measurement noise in the angular rate gyro, and $\tilde{\boldsymbol{\omega}}_w$ is the process noise in the angular rate gyro bias estimate (modeled as a zero-mean random walk about a constant). After taking the expectation of Eqn. 3.5, the continuous differential equation can be discretized and solved (approximately) via zeroth-order integration of the Taylor series expansion per

$$\begin{aligned} \hat{\boldsymbol{x}}_k = \begin{bmatrix} \boldsymbol{q}_k \\ \boldsymbol{\omega}_{bk} \end{bmatrix} &= \begin{bmatrix} \boldsymbol{q}_{k-1}\exp\{(\boldsymbol{\omega}_k - \boldsymbol{\omega}_{bk})\Delta t\} \\ 0 \end{bmatrix} \\ &\approx \begin{bmatrix} \boldsymbol{q}_{k-1} \otimes q\{(\boldsymbol{\omega}_k - \boldsymbol{\omega}_{bk})\Delta t\} \\ 0 \end{bmatrix} \end{aligned} \tag{3.6}$$

where $\otimes$ again denotes the quaternion product and

$$q\{\cdot\} = \begin{bmatrix} cos(\frac{\phi}{2}) \\ \hat{\boldsymbol{p}}sin(\frac{\phi}{2}) \end{bmatrix} \tag{3.7}$$

in which the (half) rotation angle $\phi = (\boldsymbol{\omega}_k - \boldsymbol{\omega}_{bk})\Delta t$ and the rotation axis $\hat{\boldsymbol{p}} = \boldsymbol{\omega}_k / \|\boldsymbol{\omega}_k\|$. While the process described by $f(\cdot)$ is nonlinear, it can still be used to predict the current estimated state and using the Jacobian of the process model

$$\boldsymbol{F} = \frac{\partial}{\partial \hat{\boldsymbol{x}}} f(\hat{\boldsymbol{x}}, \boldsymbol{u}) \tag{3.8}$$

to propagate the error covariance matrix to the current time step per

$$\hat{\boldsymbol{P}}_k = \boldsymbol{F}\hat{\boldsymbol{P}}_{k-1}\boldsymbol{F}^T + \boldsymbol{Q} \tag{3.9}$$

This is the premise of the EKF where the process model is linearized about the current state trajectory, which is generally accurate but inherently suboptimal.

Step 2: Update

During the update step, additional currently available information (such as that from the accelerometer and/or magnetometer) is fused with the estimated state, yielding more accurate estimations than provided in Step 1. The relationship between a measurement provided by one of the sensors in the IMU (e.g., accelerometer or magnetometer) and the estimated state can be described as

$$\hat{\boldsymbol{y}} = h(\hat{\boldsymbol{x}}) + \tilde{\boldsymbol{v}} \tag{3.10}$$

where $\hat{\boldsymbol{y}}$ is the estimated measurement provided by the measurement model, $h(\cdot)$, which is polluted by (zero-mean, Gaussian, uncorrelated) measurement noise, $\tilde{\boldsymbol{v}}$. Accordingly, the measurement model for the accelerometer is

$$\hat{\boldsymbol{y}}_{a,k} = \boldsymbol{R}\{\hat{\boldsymbol{q}}_k\}\hat{\boldsymbol{g}} + \tilde{\boldsymbol{a}}_n \tag{3.11}$$

where $\boldsymbol{R}\{\hat{\boldsymbol{q}}\}$ is the direction cosine matrix representing the sensor orientation, $\hat{\boldsymbol{g}}$ is the expected direction of gravity, and $\tilde{\boldsymbol{a}}_n$ is the accelerometer measurement noise. Similarly, the measurement model for the magnetometer is

$$\hat{\boldsymbol{y}}_{m,k} = \boldsymbol{R}\{\hat{\boldsymbol{q}}_k\}\hat{\boldsymbol{b}} + \tilde{\boldsymbol{m}}_n \tag{3.12}$$

where $\hat{\boldsymbol{b}}$ is the expected direction of the magnetic field (e.g., magnetic North) and $\tilde{\boldsymbol{m}}_n$ is the magnetometer noise.

After taking the expectations of both measurement models, the innovation, $\boldsymbol{\mathcal{I}}$,

between the actual measurement and the estimated measurement becomes

$$\boldsymbol{\mathcal{I}}_k = \boldsymbol{y}_k - \hat{\boldsymbol{y}}_k = \begin{bmatrix} \boldsymbol{y}_{a,k} - \hat{\boldsymbol{y}}_{a,k} \\ \boldsymbol{y}_{m,k} - \hat{\boldsymbol{y}}_{m,k} \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_{m,k} - \boldsymbol{R}\{\hat{\boldsymbol{q}}_k\}\hat{\boldsymbol{g}} \\ \boldsymbol{m}_{m,k} - \boldsymbol{R}\{\hat{\boldsymbol{q}}_k\}\hat{\boldsymbol{b}} \end{bmatrix} \tag{3.13}$$

where $\boldsymbol{a}_{m,k}$ is the normalized accelerometer measurement and $\boldsymbol{m}_{m,k}$ is the normalized magnetometer measurement. The uncertainty (covariance) associated with the innovation, $\boldsymbol{S}$, is defined as

$$\boldsymbol{S}_k = \boldsymbol{H}_k \hat{\boldsymbol{P}}_k {\boldsymbol{H}_k}^T + \boldsymbol{R} \tag{3.14}$$

where $\boldsymbol{R}$ is the measurement noise associated with $\tilde{\boldsymbol{v}}$ from Eqn. 3.10 and $\boldsymbol{H}$ is the Jacobian of the measurement models, $h(\cdot)$ given by

$$\begin{aligned} \boldsymbol{H} &= \frac{\partial}{\partial \hat{\boldsymbol{x}}} h(\hat{\boldsymbol{x}}) \\ &= 2 \begin{bmatrix} q_0\hat{\boldsymbol{g}} + \hat{\boldsymbol{g}} \times \boldsymbol{p} & \boldsymbol{p}^T\hat{\boldsymbol{g}}\boldsymbol{I} + \boldsymbol{p}\hat{\boldsymbol{g}}^T - \hat{\boldsymbol{g}}\boldsymbol{p}^T + q_0\hat{\boldsymbol{g}}^\times \\ q_0\hat{\boldsymbol{b}} + \hat{\boldsymbol{b}} \times \boldsymbol{p} & \boldsymbol{p}^T\hat{\boldsymbol{b}}\boldsymbol{I} + \boldsymbol{p}\hat{\boldsymbol{b}}^T - \hat{\boldsymbol{b}}\boldsymbol{p}^T + q_0\hat{\boldsymbol{b}}^\times \end{bmatrix} \end{aligned} \tag{3.15}$$

Here, q_0 is the scalar component of the quaternion, $\boldsymbol{p}$ is the vector component of the quaternion, $\hat{\boldsymbol{g}}$ is the expected direction of gravity, $\hat{\boldsymbol{b}}$ is the expected direction of magnetic North, and $\boldsymbol{I}$ is the identity matrix. For a thorough derivation of this result, see [98]. The Jacobian, $\boldsymbol{H}$, transforms the error covariance from the state space to the measurement space so the measurement noise can be appropriately added to yield a more accurate estimate of the uncertainty associated with the innovation. The Kalman filter gain is therefore defined as

$$\boldsymbol{K}_k = \hat{\boldsymbol{P}}_k {\boldsymbol{H}_k}^T {\boldsymbol{S}_k}^{-1} \tag{3.16}$$

which is used to scale the correction provided by the innovation by taking into account the uncertainty associated with both the estimated state and the measurements themselves per

$$\Delta\hat{\boldsymbol{x}}_k = \boldsymbol{K}_k \boldsymbol{\mathcal{I}}_k \tag{3.17}$$

which is incorporated into the current estimate appropriately (multiplicative or additive). The error covariance matrix is likewise updated per

$$\hat{\boldsymbol{P}}_k = (\boldsymbol{I} - \boldsymbol{K}_k \boldsymbol{H}_k)\hat{\boldsymbol{P}}_k \tag{3.18}$$

where $\boldsymbol{I}$ is an identity matrix.

This concludes the background on Kalman filtering for linear and nonlinear dynamic systems. The methods section begins by describing the Error-State Kalman Filter (ESKF) formulation by building from the descriptions above.

3.2.2 Methods

ESKF Process Model

The model herein is drawn largely from [91] and [97]. The Error-State Kalman Filter (ESKF) defines the true state of the system, $\boldsymbol{x}_t$, as the combination of a nominal-state, $\boldsymbol{x}_n$, and an error-state, $\boldsymbol{\delta x}$. The nominal-state is described by the same dynamics in Eqn. 3.5 while ignoring the process noise. The error-state dynamics are derived from the nominal-state by applying a small perturbation, $\boldsymbol{\delta x}$, about the nominal-state per

$$\dot{\boldsymbol{x}} + \dot{\boldsymbol{\delta x}} = f(\boldsymbol{x} + \boldsymbol{\delta x}) \tag{3.19}$$

The Taylor series expansion about $\boldsymbol{x}$ is

$$\dot{\boldsymbol{x}} + \dot{\boldsymbol{\delta x}} = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})\boldsymbol{\delta x} + \mathcal{O}(\boldsymbol{x}, \boldsymbol{\delta x}) \tag{3.20}$$

Neglecting higher order terms and observing $\dot{\boldsymbol{x}} = f(\boldsymbol{x})$ yields

$$\dot{\boldsymbol{\delta x}} = \boldsymbol{F}\delta\boldsymbol{x} \tag{3.21}$$

where $\boldsymbol{F} = \nabla f(\boldsymbol{x})$. Discretizing this differential equation and adding the appropriate process noise yields

$$\boldsymbol{\delta x}_k = (\boldsymbol{I} + \boldsymbol{F}_k \boldsymbol{\Delta t})\boldsymbol{\delta x}_{k-1} + \tilde{\boldsymbol{i}} \tag{3.22}$$

where $\boldsymbol{I}$ is an identity matrix and $\tilde{\boldsymbol{i}}$ is the process noise. Combining this derivation with Eqn. 3.5 yields the following continuous error-state process model

$$\boldsymbol{\delta x} = \begin{bmatrix} \boldsymbol{\delta\theta} \\ \boldsymbol{\delta\omega}_b \end{bmatrix} = \begin{bmatrix} -\boldsymbol{u}^{\times}\boldsymbol{\delta\theta} - \boldsymbol{\delta\omega}_b - \tilde{\boldsymbol{\omega}}_n \\ \tilde{\boldsymbol{\omega}}_w \end{bmatrix} \tag{3.23}$$

where $\boldsymbol{u} = \boldsymbol{\omega}_m - \boldsymbol{\omega}_b$, $\boldsymbol{\delta\theta}$ is the quaternion angle error, and $\boldsymbol{\delta\omega}_b$ is the angular rate gyro bias error, $\tilde{\boldsymbol{\omega}}_n$ is the process noise associated with $\boldsymbol{\delta\theta}$, and $\tilde{\boldsymbol{\omega}}_w$ is the process noise associated with $\boldsymbol{\delta\omega}_b$. The notation, $\boldsymbol{u}^{\times}$, denotes the skew symmetric form of the vector. Discretizing the error-state process model yields

$$\boldsymbol{\delta x}_k = \begin{bmatrix} \boldsymbol{R}\{\boldsymbol{u}_k\Delta t\}^T & -\Delta t\boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta\theta}_k \\ \boldsymbol{\delta\omega}_{b,k} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\theta}_{i,k} \\ \boldsymbol{\omega}_{i,k} \end{bmatrix} \tag{3.24}$$

where $\boldsymbol{\theta}_{i,k}$ and $\boldsymbol{\omega}_{i,k}$ are random impulses applied to the orientation and angular rate gyro bias estimates modeled by the Gaussian processes $\tilde{\boldsymbol{\omega}}_n$ and $\boldsymbol{\delta\omega}_b$. For a thorough derivation of this result, see [97].

Robust ESKF Measurement Models

The same measurement models for the accelerometer and magnetometer described in Eqns. 3.11 and 3.12 are used below. However, a methodology is needed to determine when the measurements are valid in order to avoid inappropriate corrections that may cause the filter to diverge. This section describes the methodology, illustrated by the decision tree shown in Fig. 3.1 that renders filter robust to inappropriate updates.

In Fig. 3.1, the blue boxes relate to the accelerometer data and the green boxes relate to the magnetometer data. The decision tree starts by determining whether the sensor is static based upon the acceleration and angular velocity data from the IMU. If the sensor is static (i.e., stationary), then the accelerometer measurement yields a valid estimate of the direction of gravity. If the sensor is not static, then the accelerometer measurement is polluted by the additional acceleration experienced by the rigid body to which the sensor is attached.

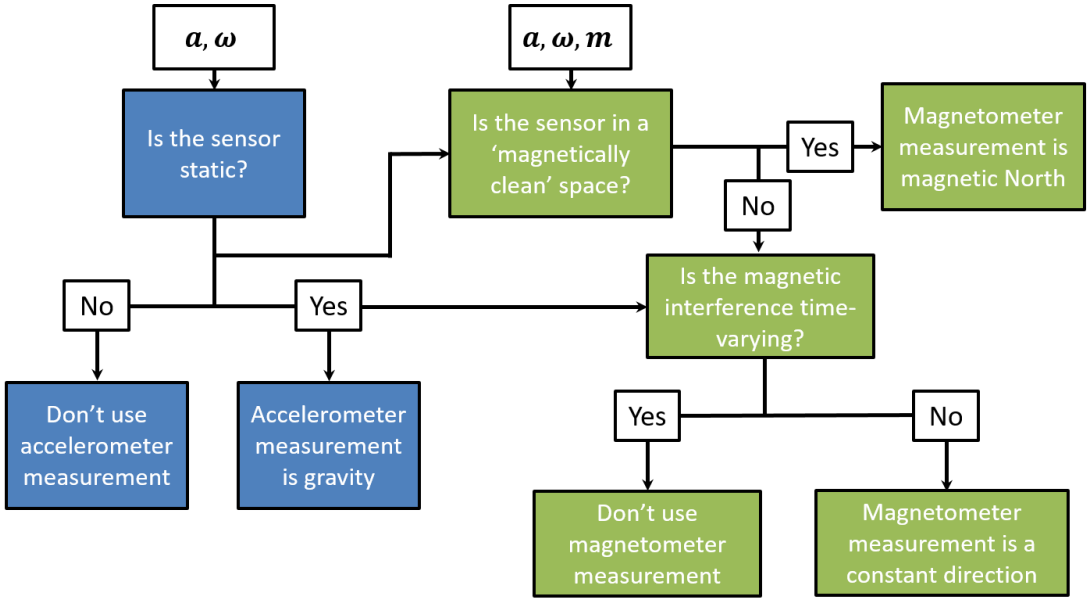


Figure 3.1: Decision tree governing when to trust or not trust the additional measurements from the accelerometer and magnetometer.

The decision tree continues by next determining if the sensor is in a 'magnetically clean' environment, and it requires the data from all three sensors (acceleration, angular velocity, and magnetometer). One criterion determines if magnetic interference is present when the sensor is static and the second criterion determines if that is the case when the sensor is not static. If no magnetic interference is detected, the magnetometer measurement yields a valid estimation of the direction of magnetic North. Beyond that, an additional and novel criterion is introduced. If magnetic interference is detected when the sensor is static, there is still an opportunity to use the magnetometer data by determining if the magnetic interference is largely constant or time-varying. If it is largely constant, then the magnetometer data is providing a constant measured direction, albeit not in the direction of magnetic North. In this instance, the magnetometer data provides an opportunity to correct for local integration drift error about the vertical axis, in which case the expected direction of the magnetic field remains the previous time step's magnetometer measurement (instead of magnetic North).

Accelerometer Criteria

The criteria for determining whether the IMU is static leverages the triangle inequality, i.e.,

$$\|\boldsymbol{x}+\boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\| \tag{3.25}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are two vectors with the same number of elements. Knowing that the accelerometer is measuring both the acceleration of the rigid body ($\boldsymbol{a}_{body}$) and the acceleration due to gravity ($\boldsymbol{g}$), Eqn. 3.25 yields

$$\|\boldsymbol{R}\boldsymbol{a}_{body}+\boldsymbol{g}\| \leqslant \|\boldsymbol{R}\boldsymbol{a}_{body}\| + \|\boldsymbol{g}\| \tag{3.26}$$

where $\boldsymbol{R}$ is the direction cosine matrix relating the IMU frame to the world frame. This can be further simplified to

$$\|\boldsymbol{a}_m\| \leqslant \|\boldsymbol{a}_{body}\| + \|\boldsymbol{g}\| \tag{3.27}$$

where $\boldsymbol{a}_m$ is the measured acceleration provided by the IMU and the magnitude of $\boldsymbol{a}_{body}$ is independent of reference frame. Clearly, if the rigid body to which the IMU is attached is not accelerating ($\|\boldsymbol{a}_{body}\| = 0$), then Eqn. 3.27 reduces to the equality

$$\|\boldsymbol{a}_m\| = \|\boldsymbol{g}\| \tag{3.28}$$

This can be relaxed slightly to account for measurement noise per

$$\big|\, \|\boldsymbol{a}_m\| - \|\boldsymbol{g}\| \,\big| \leqslant \epsilon_a \tag{3.29}$$

for which Sabatini [90] successfully uses a value of $\epsilon_a = 0.2\ m/s^2$. This represents the first of two criteria (Criterion 1) which must be satisfied simultaneously in order to determine if the accelerometer data can be used to estimate gravity.

However, there are other scenarios in which Eqn. 3.27 reduces to an equality. Consider Fig. 3.2 in which the rigid body acceleration has components in the opposite direction of ($\boldsymbol{a}_{\boldsymbol{v}}$) and perpendicular to ($\boldsymbol{a}_{\boldsymbol{h}}$) gravity.

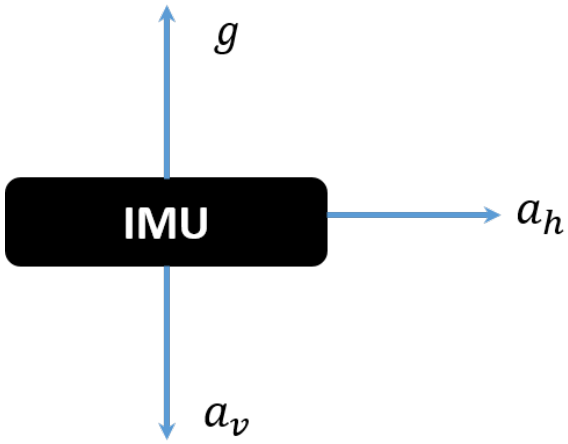


Figure 3.2: An IMU illustrated with the measured direction of gravity, $\boldsymbol{g}$, and hypothetical rigid body accelerations in the vertical direction, $\boldsymbol{a_v}$, (parallel to and in the opposite direction of measured gravity) and horizontal direction, $\boldsymbol{a_h}$ (perpendicular to gravity).

The magnitude of this measured acceleration of the IMU becomes

$$\|\boldsymbol{a}_m\| = \sqrt{(\boldsymbol{g} - \boldsymbol{a_v})^2 + \boldsymbol{a_h^2}} \tag{3.30}$$

If the magnitude of the measured acceleration is assumed to be equal to gravity, Eqn. 3.30 yields

$$\boldsymbol{a_v}(2\boldsymbol{g} - \boldsymbol{a_v}) = \boldsymbol{a_h^2} \tag{3.31}$$

which admits three solutions. The first is the trivial solution (the vertical and horizontal components of the acceleration are zero), which replicates the first scenario described above. A second solution arises when the horizontal component is zero and the vertical component equals $2\boldsymbol{g}$. The third solution arises when $\boldsymbol{a_h} = \sqrt{2\boldsymbol{g}\boldsymbol{a_v} - \boldsymbol{a_v^2}}$ for $\boldsymbol{a_v} \in [0, 2\boldsymbol{g}]$ (note the sign of $\boldsymbol{a_v}$ flips when $\boldsymbol{a_v} > 2\boldsymbol{g}$ in which case the vertical acceleration aligns with measured gravity). With respect to the two nontrivial solutions, it is highly unlikely during human movement that these accelerations will result from purely linear acceleration; that is, some proportion of the acceleration will arise from the angular kinematics as evidenced by nonzero measured angular

rate. Therefore, a second criterion (Criterion 2) considers the magnitude of the angular velocity using the Mahalanobis distance [102]. The Mahalanobis distance, D, is the generalized distance from any sample to a distribution composed of n normally distributed variables per

$$D^2 = (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \tag{3.32}$$

where $\boldsymbol{x}$ denotes the sample, $\boldsymbol{\mu}$ is the mean of the distribution, and $\boldsymbol{\Sigma}$ is the covariance of the distribution. A χ^2 test with n degrees of freedom can be used to determine if an angular velocity sample is from the normal distribution characterized by data collected during a static period at the beginning of a trial [103]. However, the distributions for the angular velocity magnitudes are necessarily positively skewed (and greater than 0), which is remedied with a natural log transformation on the angular velocity data to yield

$$\boldsymbol{x} = ln(|\boldsymbol{\omega}_m|) \tag{3.33}$$

which is used in Eqn. 3.32. Thus, when both Criterion 1 and Criterion 2 are met, the IMU is considered to be static.

Magnetometer Criteria

Determining whether there is magnetic interference present in the local environment of the IMU requires different criteria depending on if the sensor is static or not. If the sensor is static, the angle between the accelerometer measurement and magnetometer measurement is calculated per

$$\theta = \left| \cos^{-1} \left(\frac{\boldsymbol{m}_m \cdot \boldsymbol{a}_m}{\|\boldsymbol{m}_m\| \, \|\boldsymbol{a}_m\|} \right) \right| \tag{3.34}$$

which is then compared to a distribution characterized by data collected during a static period at the beginning of a trial via the squared Mahalanobis distance and a χ^2 test with a single degree of freedom. If the sensor is not static, then the yaw angular velocity is calculated from the magnetometer data and compared to that derived from the estimated state and measured angular velocity. The yaw angular

velocity magnitude from the magnetometer data ($\dot{\psi}_m$) is calculated via a simple numerical derivative per

$$\dot{\psi}_m = \left| \frac{1}{\Delta t} \cos^{-1} \left(\frac{m_k \cdot m_{k-1}}{\|m_k\| \, \|m_{k-1}\|} \right) \right| \tag{3.35}$$

The yaw angular velocity magnitude ($\dot{\psi}_\omega$) derived from the angular velocity is

$$\dot{\psi}_\omega = \left| -\frac{\sin(\phi)}{\cos(\theta)} \omega_y + \frac{\cos(\phi)}{\cos(\theta)} \omega_z \right| \tag{3.36}$$

where ϕ and θ are Euler angles representing the current estimated state (i.e., estimated quaternion converted to Euler angles) and ω_y and ω_z are measured angular velocity components. See Appendix A for a derivation of Eqn. 3.36. Similar to Eqn. 3.29, the magnitude of the difference between these two estimated yaw rates is assumed to be below some threshold, e.g.,

$$|\dot{\psi_m} - \dot{\psi}_\omega| \leqslant \epsilon_m \tag{3.37}$$

where $\epsilon_m = 10\,^\circ/s$, which is the RMS of the estimated angular velocity calculated from the (noisy) magnetometer data collected during a known static period. If it is determined that the sensor is in a 'magnetically clean' environment, then the magnetometer measurement yields a reliable estimate of magnetic North and is used accordingly.

If magnetic interference is detected, the next step is to determine if the interference is largely constant or time-varying. While the magnetometer field around the IMU varies spatially, it does not necessarily vary in time. Thus, if the IMU is static and the magnetic field is time-invariant, it then yields an estimate of a constant direction that may not be magnetic North. The constant direction can be used to correct for local drift error in orientation about the vertical axis (i.e., yaw). This is achieved by calculating the magnitude of the angular velocity of the sensor from the magnetometer measurements using Eqn. 3.35 and testing the squared Mahalanobis distance from that sample to the magnetometer-estimated angular velocity distribu-

tion characterized by the data collected during a static period. If it is determined the magnetic field is not time-varying, the expected direction of the magnetic field used in Eqn. 3.12 is the magnetometer measurement from the previous time step (instead of magnetic North).

Measurement Models

The measurement models (Eqs. 3.11 and 3.12) remain the same, but there are additional steps associated with Step 2 (update) to accommodate the ESKF formulation and the robust method described previously. First, the innovation $\mathcal{I}$ and the innovation covariance S follow from whichever measurements are found to be valid. For simplicity, the case when both the accelerometer and magnetometer measurements are valid is presented and analogous results follow quickly for all other cases.

The Jacobian of the measurement models becomes following

$$\begin{aligned} \boldsymbol{H}_{\delta x} &= \frac{\partial}{\partial \hat{\boldsymbol{\delta x}}} h(\hat{\boldsymbol{x}}) \\ &= \frac{\partial}{\partial \hat{\boldsymbol{x}}} h(\hat{\boldsymbol{x}}) \frac{\partial \hat{\boldsymbol{x}}}{\partial \hat{\boldsymbol{\delta x}}} \\ &= \boldsymbol{H} \boldsymbol{X}_{\delta x} \end{aligned} \tag{3.38}$$

where $\boldsymbol{H}$ is the Jacobian from Eqn. 3.15. The Jacobian of the estimated state with respect to the error-state is

$$\boldsymbol{X}_{\delta x} = \begin{bmatrix} \boldsymbol{Q}_{\delta x} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \tag{3.39}$$

where $\boldsymbol{I}$ is the identity matrix and

$$\boldsymbol{Q}_{\delta x} = \frac{1}{2} \begin{bmatrix} -q_1 & -q_2 & -q_3 \\ q_0 & -q_3 & q_2 \\ q_3 & q_0 & -q_1 \\ -q_2 & q_1 & q_0 \end{bmatrix} \tag{3.40}$$

For the full derivation of this result see [97].

Experimental Procedure

To validate the accuracy of the above method for estimating the orientation of a single IMU, data from the IMU is first time-synchronized to the encoder data from the coordinate measurement machine (CMM) depicted in Fig. 3.3. To this end, the assembly is rotated by hand about the CMM's base (white dashed line) for approximately five seconds. The angle measured by the optical encoder about the base axis is differentiated with respect to time yielding an angular velocity signal to compare with that measured by the IMU, which allows for measuring (and subsequently subtracting) the time delay between their respective angular rates. This movement is

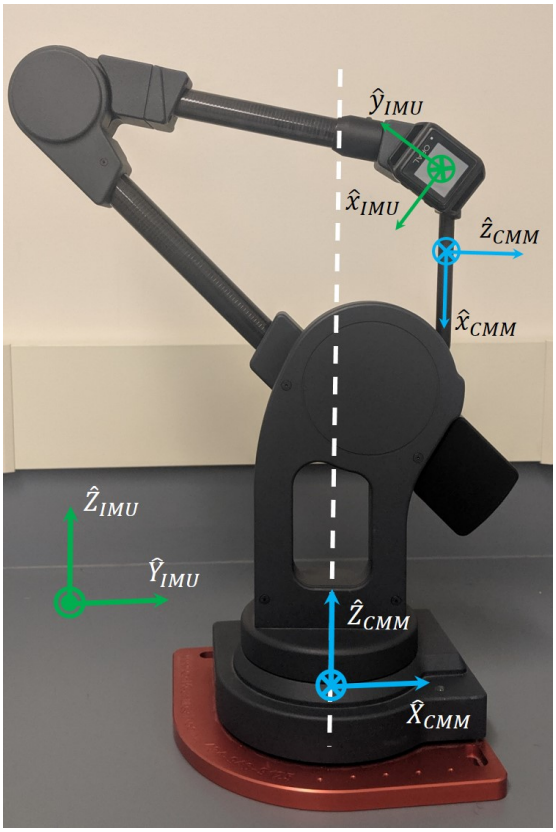


Figure 3.3: Set-up for validation of IMU-predicted orientation with the IMU attached to the CMM's end effector. Body-fixed frames and inertial frames for the IMU (green) and end effector (blue). The white dashed line denotes the rotational axis of the CMM's base.

also used to determine the orientation between the IMU's sense axes and those of the body-fixed frame of the last link in the CMM's kinematic chain (i.e., the link farthest from the base), hereafter referred to as the end effector. The $\hat{\boldsymbol{x}}_{CMM}$-axis is along the longitudinal axis of the end effector (pointing down and also aligned with the base axis in this configuration), the $\hat{\boldsymbol{y}}_{CMM}$-axis is collinear with the final joint's axis of rotation (into the page), and the $\hat{\boldsymbol{z}}_{CMM}$-axis is perpendicular to both (to the right). Since the CMM is a simple kinematic chain of single degree-of-freedom rotational joints, the orientation of the end effector is determined via sequential multiplication of direction cosine matrices (see Appendix A).

The first calculation is to provide an intermediate estimate of the $\hat{\boldsymbol{z}}_{CMM}$-axis resolved in the IMU's body-fixed frame of reference. To that end, the measured acceleration of the IMU attached to the end effector as depicted in Fig. 3.3 can be expressed as

$$\boldsymbol{a}_m = \dot{\boldsymbol{\omega}}_m \times \boldsymbol{r} + \boldsymbol{\omega}_m \times (\boldsymbol{\omega}_m \times \boldsymbol{r}) + \boldsymbol{g} \tag{3.41}$$

where $\boldsymbol{r}$ is the vector describing the position of the accelerometer embedded in the IMU relative to the origin of the CMM and $\boldsymbol{g}$ is measured gravity. Since the CMM is only rotating about vertical, the acceleration at the start of this movement (i.e., time step before the CMM starts to move) can be subtracted to remove gravity from the accelerometer measurements. Note that when the angular velocity achieves an extremum, the angular acceleration ($\dot{\boldsymbol{\omega}}_m$) is zero thereby circumventing the need for numerically differentiating the angular velocity data (and injecting another source of noise into the calculation). Thus, evaluating Eqn. 3.41 at the extremum (say n times) results in n-equations

$$\begin{aligned} \boldsymbol{a}_{m,k} &= \boldsymbol{\omega}_{m,k} \times (\boldsymbol{\omega}_{m,k} \times \boldsymbol{r}) \\ &= (\boldsymbol{\omega}_{m,k}^{\times})^2 \boldsymbol{r} \\ &= \boldsymbol{\Omega}_{m,k} \boldsymbol{r} \end{aligned} \tag{3.42}$$

where k is the time step corresponding to the k^{th} extremum in the measured angular velocity. This produces an overdetermined, linear set of n equations that can be

solved for r via

$$\boldsymbol{r} = \boldsymbol{A}/\boldsymbol{\Omega} \tag{3.43}$$

where $\boldsymbol{A}$ is a vector concatenating the accelerometer measurements, $\boldsymbol{\Omega}$ is a matrix concatenating the angular velocity measurements, and the backslash operator denotes the QR decomposition of the system of equations. Normalizing the resulting vector by its magnitude provides an intermediate estimate for the unit vector defining the direction from the IMU (accelerometer) to the CMM's origin (i.e., the $\hat{\boldsymbol{z}}_{CMM}$-axis resolved in the IMU's body-fixed frame of reference).

Next, a principal component analysis conducted on the angular velocity measured during this movement provides as its first component the $\hat{\boldsymbol{x}}_{CMM}$-axis of the end effector of the CMM resolved in the IMU's body-fixed frame of reference. One cross product ($\hat{\boldsymbol{z}}_{CMM} \times \hat{\boldsymbol{x}}_{CMM}$) yields the $\hat{\boldsymbol{y}}_{CMM}$-axis and a second cross product ($\hat{\boldsymbol{x}}_{CMM} \times \hat{\boldsymbol{y}}_{CMM}$) ensures the axes are orthonormal. Thus, the (constant) rotation matrix from the IMU frame to the CMM's end effector frame follows from

$$\boldsymbol{R}_{CMM/IMU} = \begin{bmatrix} \hat{\boldsymbol{x}}_{CMM} \\ \hat{\boldsymbol{y}}_{CMM} \\ \hat{\boldsymbol{z}}_{CMM} \end{bmatrix} \tag{3.44}$$

which is applied to the quaternions estimated from the ESKF.

Finally, the inertial frame of the CMM is different from the world frame defined by the IMU. After the IMU to CMM correction has been applied to the ESKF orientation estimates, this relationship is determined by comparing the initial orientation of the IMU and the CMM end effector at time 0, which is then applied to the quaternions estimated from the ESKF.

After the alignment movement, each of the five trials considered next lasts for approximately five minutes. During this time, the CMM end effector is actuated by hand to induce large differences in orientation with moderate dynamics. Brief pauses ($\leq$1 sec) during the data collection are also included to ensure opportunities to correct the IMU orientation.

3.2.3 Results and Discussion

Figure 3.4 illustrates representative quaternion components estimated from the IMU via the ESKF as compared to the quaternion components provided by the coordinate measurement machine (CMM) over a 30-second time interval. For an intuitive quantitative comparison, the RMS error between the true Euler angles provided by the CMM and the estimated Euler angles provided by the IMU for this trial (approximately 5 minutes) are 3.8°, 5.3°, and 5.9° for roll, pitch, and yaw, respectively. While this agreement is excellent, these errors are also the largest of the 5 trials, which have an overall RMS error of 3.4°, 3.5°, and 4.3° for roll, pitch, and yaw, respectively. These results are comparable to others reported in the literature (see, for example, [75] or [90]).

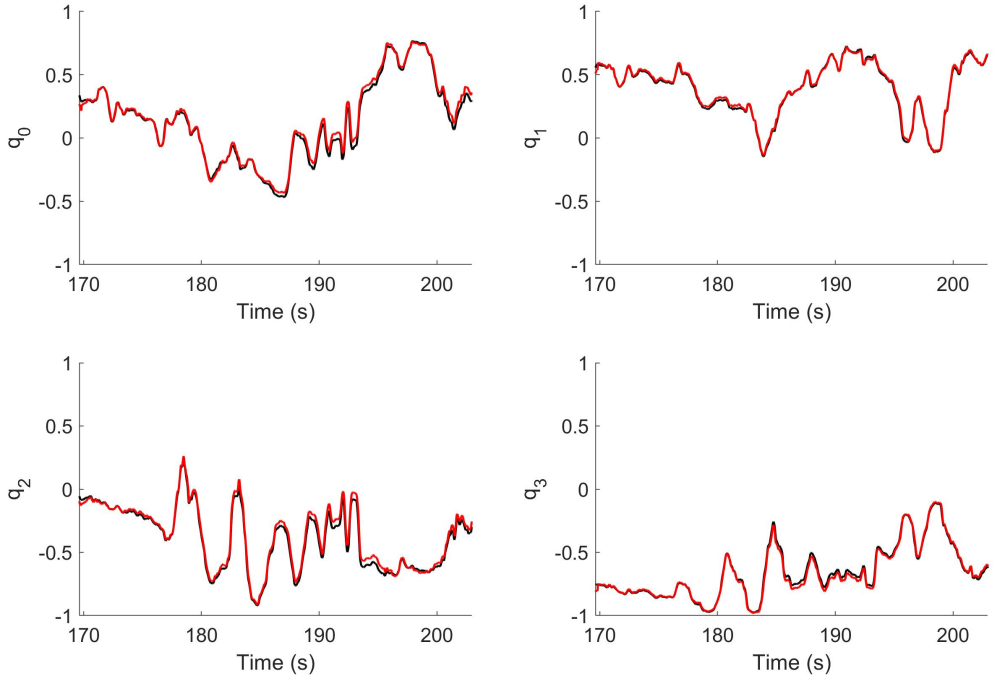


Figure 3.4: Example results comparing the CMM truth data (black) and the IMU-derived estimates (red) approximately halfway into the 5 minute trial.

These results also illustrate the effectiveness of the second stage of the magnetometer criteria described in Section 3.2.2. Specifically, the RMS errors *without* the inclusion of the second stage increase to 5.2°, 7.3°, and 7.9°, respectively, which represents an average 27% degradation. Figure 3.5 illustrates the magnitude of the magnetometer measurements from this trial (which ideally should be constant), along with a call-out corresponding to the same time period in Fig. 3.4. The large fluctuations demonstrate the challenge of using the magnetometer data to correct for integration drift error about vertical. In this case, the magnetic interference is likely

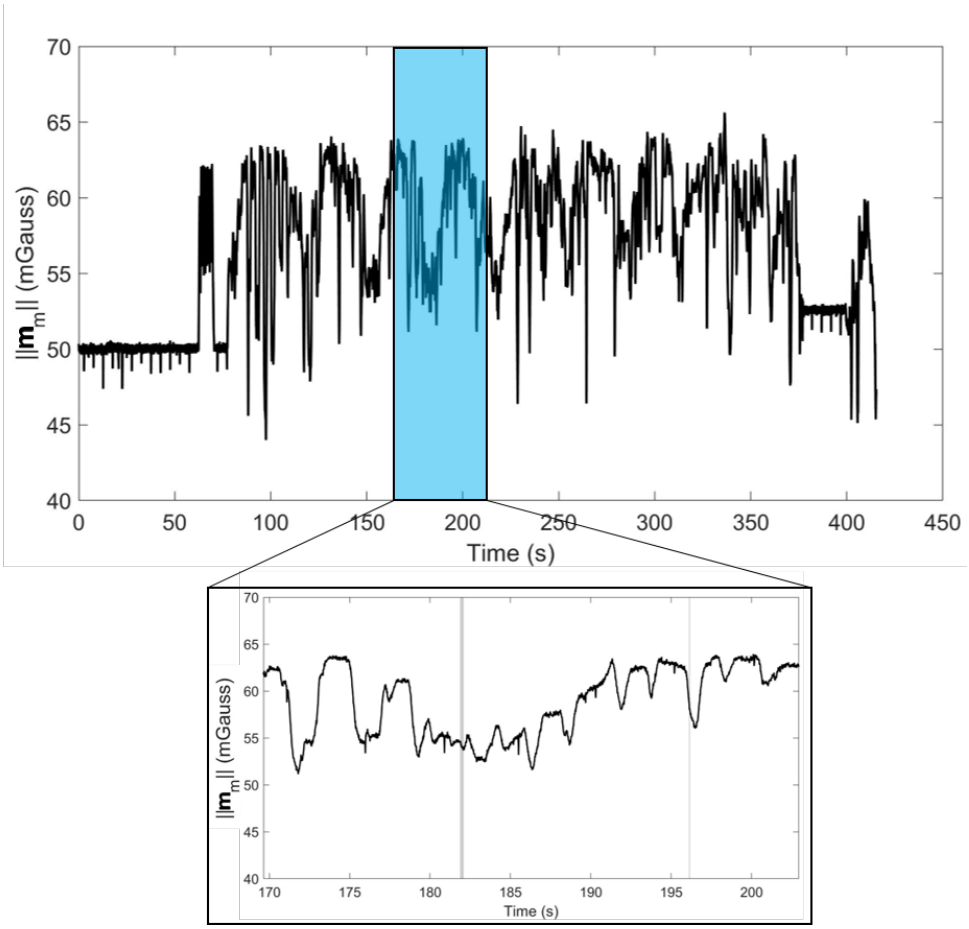


Figure 3.5: Magnetometer field magnitude with a call-out of the time period illustrated in Fig. 3.4. Grey lines denote the case when the magnetic field is largely constant and used to correct for local integration drift error.

largely due to the CMM itself, which precludes using the magnetometer data as an estimate of magnetic North. However, the grey shaded areas (see call-out) denote instances when the magnetometer data is largely constant and used to correct the estimated states.

Next, consider the results of Fig. 3.6 which show the IMU-estimated angles versus those measured by the optical encoders across all 5 testing sessions (cumulatively about 25 minutes) of combined 3D rotations.

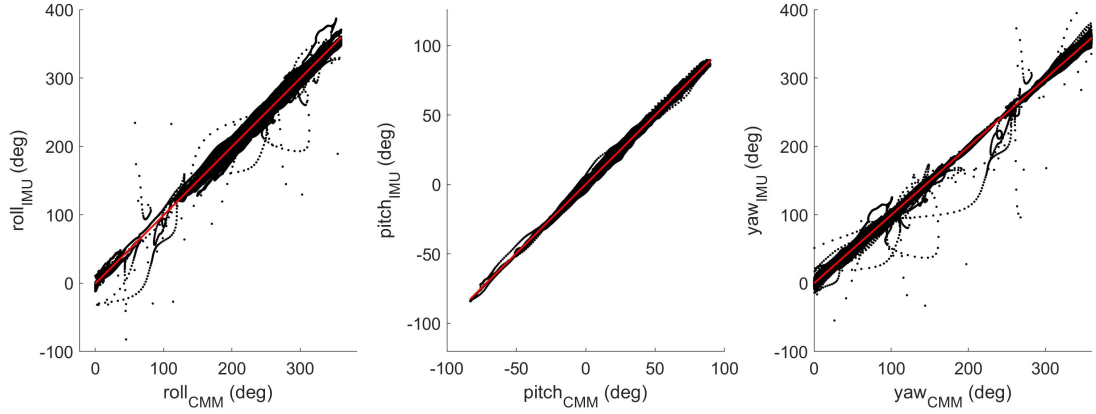


Figure 3.6: Correlation of the results for all three Euler angles from all trials. IMU-estimated angles plotted against the corresponding truth data from the CMM. The red lines are the linear fit.

The best fit lines and correlation coefficients are reported in Table 3.1. Overall, the IMU estimates exhibit excellent agreement with the CMM-provided truth data.

Table 3.1: Quantitative comparisons of Euler angles across all five trials, including RMS error, correlation (r), and slope, and y-intercept (b) of linear fit.

	RMS Error (°)	$\boldsymbol{r}$	**Slope**	$\boldsymbol{b}$ (°)
Roll	3.4	0.99	1.00	-0.77
Pitch	3.5	0.99	0.99	0.13
Yaw	4.3	0.99	1.00	-0.07

3.2.4 Conclusions

This study contributes a robust Error-State Kalman Filter (ESKF) for estimating IMU orientations that yields highly accurate results when benchmarked on a coordinate measurement machine (CMM). Included in this formulation is a novel treatment of magnetometer data when local magnetic interference is detected. Specifically, when the IMU is static and the magnetometer data is largely constant, this data is still provides a constant direction of the magnetic field that can be used to correct integration drift error about vertical. As a result, the agreement between the IMU-estimated orientation and the orientation provided by the CMM is excellent.

3.3 Study 2: Robust ESKF Validation for Two IMUs

3.3.1 Background

Leveraging the results from Study 1, this study seeks to extend the ESKF formulation from one to two sensors to then estimate the three-dimensional angles between their respective measurement frames. This is accomplished by incorporating the anatomical kinematic constraint described in Chapter 2 in which the knee analog (i.e., the coordinate measurement machine) frequently acts as a hinge. Specifically, the world frame correction, $\boldsymbol{C}_{WT/WS}$, from Chapter 2 will be incorporated into a measurement model to be used to update the orientation (Step 2) of one of the sensors. While the results in Study 1 demonstrated a strong dependence on magnetometer data for accurate estimates of *absolute* orientation for a single IMU, this study compares the accuracy of the *relative* orientation estimates between two IMUs and less dependency on magnetometer data. Specifically, the anatomical kinematic constraint (Chapter 2) and the novel magnetometer criterion (Section 3.2.2) are used as strategies to update the integration drift error between the sensors. It is hypothesized that there will be good agreement between the CMM-derived orientation (truth data) and the IMU-derived orientation since the *relative* drift about vertical between the sensors will be constrained by these two update strategies. Furthermore, a method is introduced for estimating the orientation of each sensor relative to its

respective link frame (i.e., its body-fixed frame) without relying upon the alignment movements used in Study 1 and Chapter 2. Doing so reduces yet another source of error and allows for a clearer interpretation of the efficacy of the ESKF.

3.3.2 Methods

Probabilistic Construction of a Common World Frame

As outlined in Chapter 2, a key challenge in estimating *relative* orientations of independent IMUs lies in constructing a common world frame of reference. As before, one can exploit the constraint that the hinge (medial-lateral) axes, $\hat{\boldsymbol{n}}_S$ and $\hat{\boldsymbol{n}}_T$, should be collinear in a common world frame during the time intervals when the knee can reliably be simplified to a hinge joint. Whenever this hinge criterion is satisfied, one of the IMUs' hinge axes, say $\hat{\boldsymbol{n}}_S$, serves as a reference direction for the other, say $\hat{\boldsymbol{n}}_T$. Specifically, the shank IMU's hinge axis is resolved into its world frame via

$$\hat{\boldsymbol{n}}_{ST} = \boldsymbol{R}\{\hat{\boldsymbol{q}}_{S,k}\}\hat{\boldsymbol{n}}_S \tag{3.45}$$

where k is the k^{th} time step and this vector is used in the measurement model

$$\hat{\boldsymbol{y}}_{n_{ST},k} = \boldsymbol{R}\{\hat{\boldsymbol{q}}_{T,k}\}\hat{\boldsymbol{n}}_{ST} + \tilde{\boldsymbol{\omega}}_{n_{ST}} \tag{3.46}$$

where $\boldsymbol{R}\{\hat{\boldsymbol{q}}_T\}$ is the direction cosine matrix representing the thigh sensor's orientation, $\hat{\boldsymbol{n}}_{ST}$ is the expected direction of the shank's hinge axis, and $\tilde{\boldsymbol{\omega}}_{n_{ST}}$ is the measurement noise associated with estimating $\hat{\boldsymbol{n}}_S$ and $\hat{\boldsymbol{n}}_T$ via a functional alignment movement. Specifically, this functional alignment movement consists of exercising the thigh and shank links purely about the CMM flexion-extension axis (while locking the other two rotational degrees of freedom). Following the protocol from Chapter 2, the resulting angular velocity $\boldsymbol{\omega}_S$ and $\boldsymbol{\omega}_T$ measured by the shank and thigh, respectively, define a unit vector aligned with the medial-lateral (hinge) axis measured in $\mathcal{F}_{\mathbf{S}}$ and $\mathcal{F}_{\mathbf{T}}$, respectively. The Jacobian for this measurement model becomes

$$\boldsymbol{H} = 2\begin{bmatrix} q_0\hat{\boldsymbol{n}}_{ST} + \hat{\boldsymbol{n}}_{ST} \times \boldsymbol{p} & \boldsymbol{p}^T\hat{\boldsymbol{n}}_{ST}\boldsymbol{I} + \boldsymbol{p}\hat{\boldsymbol{n}}_{ST}^T - \hat{\boldsymbol{n}}_{ST}\boldsymbol{p}^T + q_0\hat{\boldsymbol{n}}_{ST}^{\times} \end{bmatrix} \tag{3.47}$$

where q_0 is the scalar component and $\boldsymbol{p}$ is the vector component of the thigh IMU's estimated quaternion. Next, the criteria for determining when the hinge assumption is valid are redefined.

Robust ESKF Measurement Models

The methodology described in Study 1 to determine when the measurements are valid needs to be adapted to incorporate the anatomical kinematic constraint. Accordingly, the former decision tree in Fig. 3.1 (that renders the ESKF robust) is slightly altered to form the new decision tree shown in Fig. 3.7. As before, the blue boxes relate to the accelerometer data, the green boxes relate to the magnetometer data, and the purple boxes relate to the anatomical kinematic constraint.

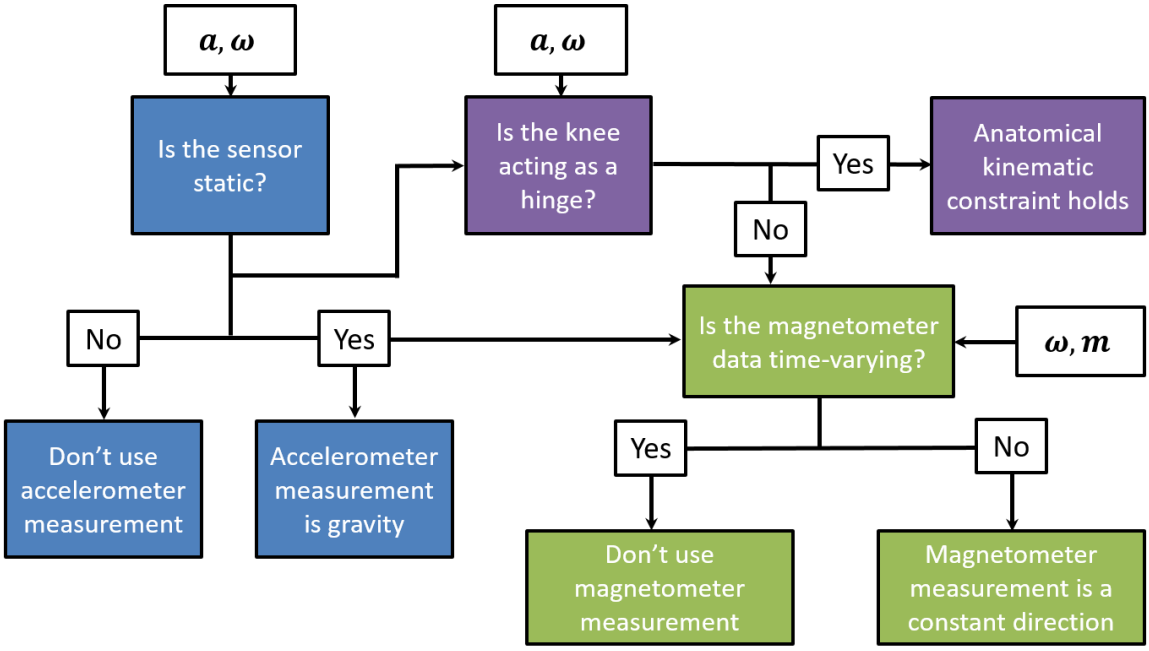


Figure 3.7: Modification of previous decision tree governing when to trust or not trust the measurements from the accelerometer, magnetometer, and anatomical kinematic constraint. Note the inclusion of a test of to determine when the knee acts as a hinge and the anatomical kinematic constraint (purple boxes).

The decision tree starts by determining whether the sensor is static based on the acceleration and angular velocity data from the IMU. Then, the decision tree

continues (using the result of that first stage) by next determining if the anatomical kinematic constraint holds using either the accelerometer or angular velocity data. If the knee analog is reliably acting as a hinge, the hinge axes from each sensor should be the same when resolved in the world frame. Finally, if the sensor is static and the knee analog is not acting as a hinge, then the decision tree determines if the magnetometer data is largely constant or time-varying. If it is largely constant, then the magnetometer is providing a constant measured direction (likely not in the direction of magnetic North).

Anatomical Kinematic Constraint Criteria

The criteria for determining when the knee can reliably be simplified to a hinge (Eqns. 2.11 and 2.13 from Chapter 2) need to be redefined, namely: Case 1a) when the knee is approximately stationary, and Case 1b) when the knee is rotating. Determining if the knee is stationary or moving is simple Boolean logic using the results from the robust decision tree described in Section 3.2.2. Specifically, the knee is considered static when both sensors are independently determined to be static and the knee is considered moving when both sensors are determined to be not static. Furthermore, these criteria are modified to implement the Mahalanobis distance (Eqn. 3.32) to determine the likelihood that the criteria are satisfied instead of the thresholds used previously.

1a) Functioning as a Hinge - Stationary Case: The knee may be considered a hinge joint when the segments are stationary and the angles between the hinge axes, $\hat{\boldsymbol{n}}_S$ and $\hat{\boldsymbol{n}}_T$, and directions of gravity measured in their respective sensor fixed frames, $\hat{\boldsymbol{g}}_S$ and $\hat{\boldsymbol{g}}_T$, are approximately the same. In such instances, the knee functions as a hinge with zero hinge rotation rate. The measurement becomes

$$x = \left| \cos^{-1}\left(\hat{\boldsymbol{n}}_S \cdot \hat{\boldsymbol{g}}_S\right) - \cos^{-1}\left(\hat{\boldsymbol{n}}_T \cdot \hat{\boldsymbol{g}}_T\right) \right| \tag{3.48}$$

in which $\hat{\boldsymbol{n}}_S$, $\hat{\boldsymbol{n}}_T$, $\hat{\boldsymbol{g}}_S$, and $\hat{\boldsymbol{g}}_T$ are all unit vectors. The mean (μ) and variance (Σ) are determined from the data from a static pose.

1b) Functioning as a Hinge - Rotating Case: When the angular velocities of the

shank and the thigh are predominantly aligned with the hinge axis defined by each segment respectively. The measurement becomes

$$x = \text{mean}\left\{ \frac{|\boldsymbol{\omega}_S \cdot \hat{\boldsymbol{n}}_S|}{|\vec{\boldsymbol{\omega}}_S|}, \frac{|\boldsymbol{\omega}_T \cdot \hat{\boldsymbol{n}}_T|}{|\vec{\boldsymbol{\omega}}_T|} \right\} \tag{3.49}$$

in which $\boldsymbol{\omega}_S$ and $\boldsymbol{\omega}_T$ are the angular velocities measured at the shank and thigh, respectively. The mean (μ) and variance (Σ) are determined from the data from the functional alignment movement used to calculate $\hat{\boldsymbol{n}}_S$ and $\hat{\boldsymbol{n}}_T$.

Body-Fixed Frame Alignment

Next, the relationships between the IMU (body-fixed) frames and the frames defined by their respective links is determined via optimization rather than the functional alignment movements used previously. The motivation for the optimization is to study the accuracy of the estimation methodology independent of potential misalignment errors between the body-fixed frames.

Consider Eqn. 2.9 from Chapter 2 expressed in terms of quaternions instead of direction cosine matrices,

$$\boldsymbol{q}(t)_{AT/AS} = \boldsymbol{q}_{AT/T} \otimes \boldsymbol{q}(t)_{T/WT} \otimes \boldsymbol{q}_{\boldsymbol{C}}(t)_{WT/WS} \otimes \boldsymbol{q}(t)_{WS/S} \otimes \boldsymbol{q}_{S/AS} \tag{3.50}$$

where $\boldsymbol{q}(t)_{AT/AS}$ is the time-varying estimated orientation of the anatomical shank frame relative to the anatomical thigh frame. Let's say there exists some misalignment between the anatomical frames defined by the IMUs and the true anatomical frames of their respective CMM links. Equation 3.50 can be rewritten to include these potential misalignment errors per

$$\boldsymbol{q}(t)_{AT'/AS'} = \boldsymbol{q}_{AT'/AT} \otimes \boldsymbol{q}(t)_{AT/AS} \otimes \boldsymbol{q}_{AS/AS'} \tag{3.51}$$

where AT' and AS' denote the true anatomical frames for the thigh and shank links, respectively. Assuming the IMUs are affixed to the links such that there is no relative movement between them, these misalignments ($\boldsymbol{q}_{AT'/AT}$ and $\boldsymbol{q}_{AS/AS'}$) are constant. Since the *relative* orientations provided by the CMM and estimated by the IMUs

are known, Eqn. 3.51 can be re-expressed as a linear, homogeneous, overdetermined system of n equations of the form

$$\begin{aligned}
\boldsymbol{q}(t_k)_{AT'/AS'} \otimes \boldsymbol{q}^*_{AS/AS'} &= \boldsymbol{q}_{AT'/AT} \otimes \boldsymbol{q}(t_k)_{AT/AS} \\
[\boldsymbol{q}(t_k)_{AT'/AS'}]_L \, \boldsymbol{q}^*_{AS/AS'} &= [\boldsymbol{q}(t_k)_{AT/AS}]_R \, \boldsymbol{q}_{AT'/AT} \\
[\boldsymbol{q}(t_k)_{AT'/AS'}]_L \, \boldsymbol{q}^*_{AS/AS'} - [\boldsymbol{q}(t_k)_{AT/AS}]_R \, \boldsymbol{q}_{AT'/AT} &= 0 \\
\begin{bmatrix} [\boldsymbol{q}(t_k)_{AT'/AS'}]_L & -[\boldsymbol{q}(t_k)_{AT/AS}]_R \end{bmatrix} \begin{bmatrix} \boldsymbol{q}^*_{AS/AS'} \\ \boldsymbol{q}_{AT'/AT} \end{bmatrix} &= 0 \\
\boldsymbol{A}_k \boldsymbol{x}_q &= 0
\end{aligned} \tag{3.52}$$

where t_k denotes the sampled-time for the k^{th} time step, $\boldsymbol{q}^*$ denotes the conjugate of the quaternion, and $[\boldsymbol{q}]_L$ or $[\boldsymbol{q}]_R$ denote the left- or right- quaternion-product matrices per [97]

$$[\boldsymbol{q}]_L = \begin{bmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{bmatrix} \tag{3.53}$$

$$[\boldsymbol{q}]_R = \begin{bmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & q_3 & -q_2 \\ q_2 & -q_3 & q_0 & q_1 \\ q_3 & q_2 & -q_1 & q_0 \end{bmatrix} \tag{3.54}$$

Collecting all n equations (one for every time step in the data set) yields

$$\begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \\ \vdots \\ \boldsymbol{A}_n \end{bmatrix} \boldsymbol{x}_q = \boldsymbol{A}\boldsymbol{x}_q = 0 \tag{3.55}$$

To find a non-trivial solution to this system of equations, a constrained least-squares

problem is created

$$\boldsymbol{x}_q^* = \arg\max_{\boldsymbol{x}_q} \|\boldsymbol{A}\boldsymbol{x}_q\| \text{ subject to } \|\boldsymbol{x}_q\| = 1 \tag{3.56}$$

such that it can be solved by defining a Lagrangian function

$$\begin{aligned} \mathbf{L}(\boldsymbol{x}_q, \lambda) &= \|\boldsymbol{A}\boldsymbol{x}_q\| + \lambda(1 - \|\boldsymbol{x}_q\|) \\ &= \boldsymbol{x}_q^T \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x}_q + \lambda(1 - \boldsymbol{x}_q^T \boldsymbol{x}_q) \end{aligned} \tag{3.57}$$

The critical points (i.e., points where the derivatives of $\mathbf{L}(\cdot, \cdot) = 0$) correspond to nontrivial solutions.

$$\frac{\partial}{\partial \boldsymbol{x}_q}\mathbf{L}(\boldsymbol{x}_q, \lambda) = 2\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}_q - 2\lambda\boldsymbol{x}_q = 0 \tag{3.58}$$

$$\frac{\partial}{\partial \lambda}\mathbf{L}(\boldsymbol{x}_q, \lambda) = 1 - \boldsymbol{x}_q^T\boldsymbol{x}_q = 0 \tag{3.59}$$

Note that Eqn. 3.58 can be expressed as a characteristic equation, meaning $(\boldsymbol{A}^T\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x}_q = 0$, which when combined with the constraint described by Eq. 3.59

$$\begin{aligned} \|\boldsymbol{A}\boldsymbol{x}_q\| &= \boldsymbol{x}_q^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}_q \\ &= \boldsymbol{x}_q^T\lambda\boldsymbol{I}\boldsymbol{x}_q \\ &= \lambda \end{aligned} \tag{3.60}$$

reveals the solution of interest is the eigenvector corresponding to the smallest eigenvalue of $\boldsymbol{A}^T\boldsymbol{A}$. While the magnitude of $\boldsymbol{x}_q$ was constrained to 1, the actual magnitude is $\sqrt{2}$ since $\boldsymbol{x}_q$ is two stacked unit quaternions. Thus, $\boldsymbol{x}_q$ is normalized by $\sqrt{2}$ to yield the proper quaternions relating the orientation of the anatomical frames defined by the IMUs to the true anatomical frames of their respective CMM links.

Experimental Procedure

Data from the two IMUs are first time-synchronized to the encoder data from the CMM (Fig. 3.8). The assembly is rotated by hand about the CMM's base

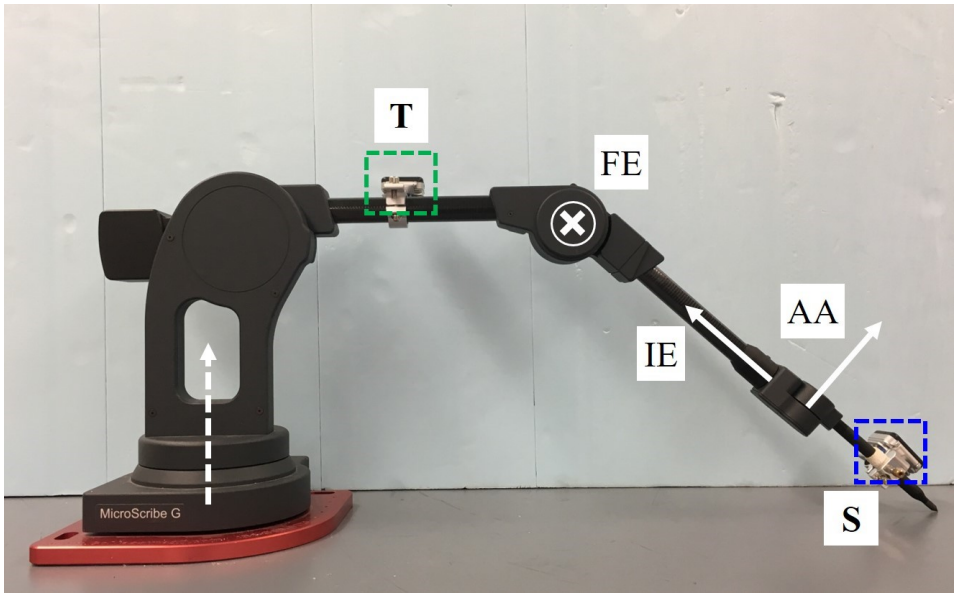


Figure 3.8: Knee analog formed by a coordinate measurement machine (CMM). Three anatomical axes for flexion-extension (FE), internal-external rotation (IE), and abduction-adduction (AA) are labeled are the corresponding rotational joints of the CMM. Two labeled IMUs are mounted to the CMM with **T** (green) analogous to a thigh-mounted IMU and **S** (blue) analogous to a shank-mounted IMU.

with the three knee axes (FE, IE, AA) locked in their neutral angular positions (i.e., zero angles). The angle measured by the optical encoder about the base axis is differentiated with respect to time yielding an angular velocity signal to compare with those measured by the thigh (green) and shank (blue) IMUs. The data from the two IMUs are already time-synchronized, and their synchronization with the data from the CMM follows from measuring (and subsequently subtracting) the time delay between their respective angular rates. Next, the two functional alignment movements are conducted to define the distributions characterizing when the knee is acting as a hinge for Case 1a) when the knee is approximately stationary and Case 1b) when the knee is rotating. First, the shank and thigh are held in a static neutral pose for about 10 seconds. The measured acceleration for each segment is the direction of gravity, which is used in Eqn. 3.48 to determine a mean and variance characterizing when the stationary knee is acting as a hinge. Next, the CMM is

exercised purely about the flexion-extension axis with the two remaining knee axes locked. The angular velocities measured during this phase are used in Eqn. 3.49 to determine a mean and variance characterizing when the rotating knee is acting as a hinge. Following this, the knee analog is exercised for approximately two minutes about all three axes to approximately mimic ranges of motion in a healthy human knee. This procedure is repeated five times.

3.3.3 Results and Discussion

Figure 3.9 illustrates representative quaternion components estimated from the IMUs via the ESKF as compared to the quaternion components provided by the coordinate measurement machine (CMM) for a 50-second time window. The quaternions describe the relative orientation of the thigh and shank links on the CMM. The

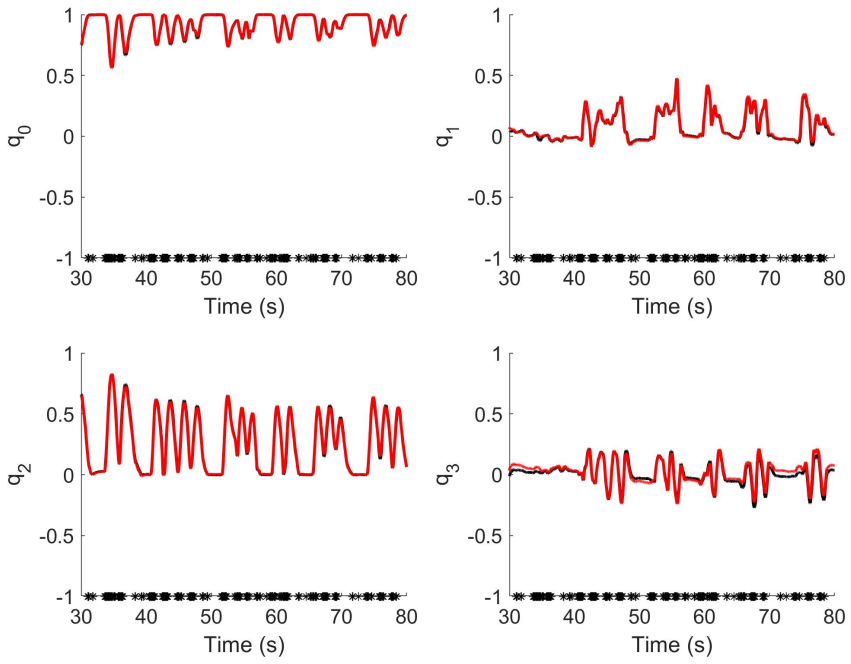


Figure 3.9: Example results comparing the CMM truth data (black) and the IMU-derived estimates (red). The black asterisks at the bottom of each graph denote time steps when the knee analog is acting predominantly as a hinge.

RMS errors between the true Euler angles provided by the CMM and the estimated Euler angles provided by the IMUs for this example trial are 3.7°, 3.0°, and 3.9° for flexion-extension, internal-external rotation, and abduction-adduction, respectively. Comparatively, the RMS errors across all 5 trials is 3.0°, 2.5°, and 5.4° for flexion-extension, internal-external rotation, and abduction-adduction, respectively. For comparison, the RMS errors from Study 1 are 3.5°, 3.4°, and 4.3° for pitch, roll, and yaw, respectively. Thus, it is reasonable that the largest RMS errors for the joint angles are those for abduction-adduction as they rely the most on the yaw estimates.

Next, consider Fig. 3.10, which shows the IMU-estimated angles versus those provided by the optical encoders for the duration of all 5 testing sessions of combined 3D rotations. The results of the best fit lines and correlation coefficients are reported in Table 3.2.

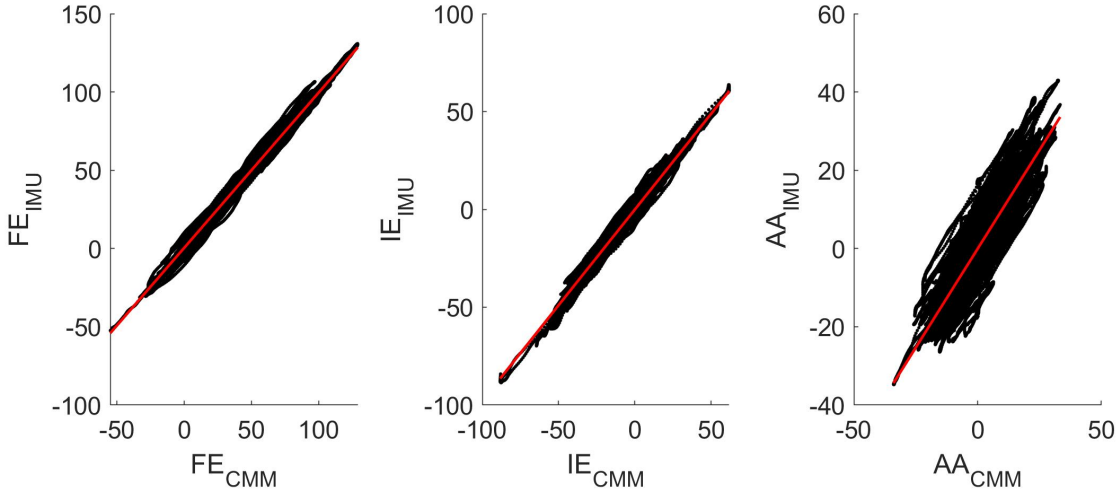


Figure 3.10: Correlation of the results for all three angles from all trials. IMU-estimated angles plotted against the corresponding truth data from the CMM. The red lines are the linear fit.

Overall, the IMU estimates for flexion-extension and internal-external rotation exhibit excellent agreement ($\boldsymbol{r} > 0.9$) with the truth data provided by the CMM whereas the estimates abduction-adduction exhibit good agreement ($\boldsymbol{r} > 0.8$). Specifically, the abduction-adduction IMU estimates exhibit larger variation resulting in

a lower correlation coefficient. The slope (1.01) suggests the variation is distributed equally about the true values measured by the CMM. Interestingly, the trial with the most frequent opportunities to use the anatomical kinematic constraint (28%) exhibited the lowest average RMS error (3.1°) and the trial with the least frequent opportunities (10%) exhibited the highest average RMS error (4.2°).

Table 3.2: Quantitative comparisons of Euler angles including RMS error, correlation (r), and slope, and y-intercept (b) of linear fit.

	RMS Error (°)	$\boldsymbol{r}$	**Slope**	$\boldsymbol{b}$ (°)
Flexion-Extension	3.0	0.99	0.99	0.46
Internal-External	2.5	0.99	0.98	-0.08
Abduction-Adduction	5.4	0.85	1.01	-0.08

For comparison to the methodology in Chapter 2, Fig. 3.11 illustrates the IMU-estimated angles plotted against those provided by the optical encoders for all 5 testing sessions of combined 3D rotations. Specifically, the orientations of each sensor are estimated using the algorithm from Study 1 for which the world frames are aligned using the deterministic treatment of the anatomical kinematic constraint.

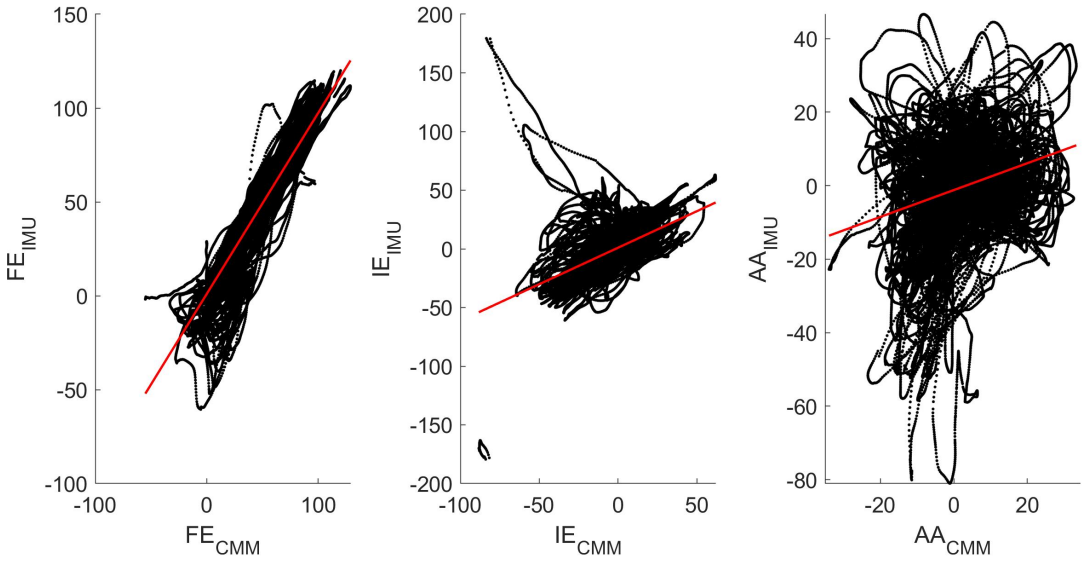


Figure 3.11: Correlation of the results for all three angles from all trials. IMU-estimated angles (using the methodology from Chapter 2) plotted against the corresponding truth data from the CMM. The red lines are the linear fit.

Overall, the IMU estimates for flexion-extension exhibit excellent agreement ($\boldsymbol{r} > 0.9$) with the truth data provided by the CMM whereas the estimates for internal-external rotation and abduction-adduction exhibit poor agreement. Table 3.3 documents the results of these analyses. Clearly, the probabilistic treatment of the anatomical kinematic constraint is superior to that of the deterministic treatment.

Table 3.3: Quantitative comparisons of Euler angles including RMS error, correlation (r), and slope, and y-intercept (b) of linear fit.

	RMS Error (°)	$\boldsymbol{r}$	**Slope**	$\boldsymbol{b}$ (°)
Flexion-Extension	11.0	0.94	0.96	0.95
Internal-External	16.1	0.53	0.62	0.96
Abduction-Adduction	13.7	0.26	0.36	-1.19

One significant difference between the experiment from Chapter 2 and the one in Study 2 is the experimental protocol. In Chapter 2, the 3D rotations experiment

consisted of exercising the CMM about all 3 axes for 10 second bouts (repeated 50 times), essentially ensuring that the hinge axis assumption is fulfilled before and after each bout. In this experiment, the CMM is exercised about all 3 axes for two minutes with sometimes limited opportunities for the hinge axis assumption to be fulfilled. While the criteria used to determine when the knee analog was acting as a hinge worked in the previous study, they were likely too strict to achieve the same effect for this (significantly more challenging) data set.

3.3.4 Conclusion

This study contributes an extension of the robust Error-State Kalman Filter (ESKF) from a single IMU from Study 1 to a pair of shank- and thigh-mounted IMUs needed to estimate 3D rotations across a knee analog. This formulation incorporates the anatomical kinematic constraint into Step 2 (update) of the shank sensor's orientation estimation to ensure that the world frames provided by the (independent) IMUs are the same. Importantly, while the *absolute* integration drift error about vertical is no longer corrected using magnetometer data and the *relative* integration drift error is corrected predominantly with the anatomical kinematic constraint. However, the novel use of the magnetometer data as a largely constant direction when the IMUs are static from Study 1 was still implemented. Results from this extended theory are benchmarked using 3D rotations using a coordinate measurement machine (CMM). Consequently, the agreement between the IMU-estimated joint angles and those measured by the embedded optical encoders is excellent, which was in part due to the new methodology for determining the orientation between the IMUs' body-fixed frames and their respective links' body-fixed frames. Finally, the (probabilistic) criteria for determining when the knee analog can conservatively and reliably be simplified to a hinge joint is successful.

3.4 Summary and Conclusions

This chapter contributes a successful extension of the measurement theory for 3D knee rotations described in Chapter 2 to a purely probabilistic theory through two steps. The first step, described as Study 1, replaces the sensor orientation estimates previously obtained using proprietary software with estimates obtained by a customized Error-State Kalman Filter (ESKF). The results exhibit excellent agreement with orientation estimates from highly precise optical encoders embedded in the CMM that serve as the ground truth. The second step, described as Study 2, further customizes the ESKF by replacing the deterministic treatment of the anatomical kinematic constraint with a probabilistic one. Compared to the method explored in Chapter 2, the extension of the method is similarly successful in its ability to estimate 3D knee rotations during extended periods of dynamic movements as will be further studied in Chapter 4.

CHAPTER 4

Application of Measurement Theory to Human Subjects and Validation with MOCAP

4.1 Introduction

This chapter seeks to extend and validate the measurement theory of the prior chapter for use on human subjects and to compare the resulting 3D joint angles to measurements from optical motion capture (MOCAP). In Chapters 2 and 3, the rigid and well-defined joints of the coordinate measurement machine (CMM) simplifies the challenge of resolving 3D joint angles. By contrast, experiments on human subjects are susceptible to additional considerations including the imprecise determination of the thigh and shank anatomical axes, the laxity of the knee, and the superimposed motions of the underlying soft tissues, among others. Although the results presented in Chapter 2 demonstrate excellent agreement with truth data, that success depends on the criteria for determining when the world frame correction is updated as well as how trustworthy that correction is. Replacing that deterministic correction with a probabilistic correction in Chapter 3 yields superior results. However, the criteria for determining when the human knee is acting as a hinge must be investigated as it will likely need to be subject-specific and perhaps movement-specific as well.

Several studies (see, for example, [104–106]) measure or simulate errors in 3D joint angles due to misalignment of the estimated anatomical frames for a body segment across motion capture techniques. This is significant for two reasons. First,

optical motion capture is susceptible to this type of error compared with truth data provided by stereoradiography/dual-plane fluoroscopy or similar techniques (see, for example, [107–111]). For this reason, the comparison between MOCAP and IMU estimates will be discussed in terms of *differences* instead of in terms of *errors*, since optical motion capture cannot reliably yield ground truth results. Second, recall the misalignment (albeit slight) observed in Chapter 2 between the "anatomical" frames of the CMM and those defined by the IMUs using the alignment movements, particularly for the abduction-adduction axes. This underscores the challenge of defining common anatomical frames across different motion capture techniques, a challenge amplified in biomechanics studies since no convention exists for defining anatomical frames via inertial sensors. Thus, the small differences in anatomical frames observed for the prior studies on a CMM will be significantly accentuated with human subjects. This challenge motivates serious consideration for determining IMU-based anatomical frames. While some misalignment between IMU-based anatomical frames and MOCAP-based anatomical frames will contribute to differences in estimated joint angles, differences will also arise from increasingly dynamic movements that induce movement of the underlying soft tissue (i.e., skin artefacts).

The overall objective of this chapter is to successfully extend the measurement theory developed in Chapter 3 to human subjects and to compare estimated joint angles to those obtained by optical motion capture. Recall from Chapter 3 that the orientation of the IMUs relative to their respective links on the CMM was determined via an optimization method instead of from the functional alignment movements used in Chapter 2. The optimization method will again be employed in this study to compare the anatomical frames determined by the IMUs to those determined by optical motion capture. Before doing so, this chapter first provides a systematic survey of the various approaches reported in the literature for defining anatomical frames of reference for inertial motion capture.

4.2 Survey of IMU Anatomical Frame Definitions

4.2.1 Motivation

In 1983, Grood and Suntay [112] proposed in their landmark publication a method of defining joint coordinate systems (with particular emphasis on the knee) such that the convention would yield rotations of clinical significance that could be easily understood by a broader audience. Following their lead several years later, the Standardization and Terminology Committee of the International Society of Biomechanics began the process of standardizing how to report kinematic data for every joint [18]. In their seminal paper, Wu and Cavanagh observed "*some uniformity in presentation will make publications easier to read and allow for the more straightforward comparison of data sets from different investigators*" [18, p. 1257]. However, these early efforts and those that followed (most notably [113, 114]) were focused solely on how to report kinematic data for *optical* motion capture systems which remain prevalent today. Given the ease with which position data is collected using optical motion capture, it is advantageous to define conventions based on position alone. In particular, the conventions for defining anatomical frames require identifying the positions of and placing markers upon anatomical landmarks as the first step.

Unfortunately, this method of defining anatomical frames does not translate directly to *inertial* motion capture, whose strength lies in its ability to measure *motion* (acceleration and angular velocity) rather than *position*. Consequently, the inertial motion capture community finds itself in a similar state as the optical motion capture community roughly 40 years ago; namely, there is no common convention for defining anatomical frames of reference using this new technology. Between advancements to IMU hardware and signal processing techniques, the accuracy of orientation estimates from IMUs have improved considerably despite the inevitable drift errors that remain. In fact, differences in joint angles between inertial and optical motion capture due to drift errors are of the same magnitude as those caused by differing anatomical frame definitions. Thus, it is increasingly important to address how

anatomical frames should be defined in order to advance inertial motion capture for the broader research community.

The objective of the literature survey below is to catalogue the different methods used to define anatomical frames of reference for inertial motion capture and to measure if the community is converging on a specific method or not. While not exhaustive, this survey certainly includes representative studies.

4.2.2 Methods

Database searches were conducted in August 2018 using Pubmed (1781-2018), Web of Science (1900-2018), and Scopus (1788-2018) with the following search terms:

- (inertial sensor OR wearable sensor OR accelerometer OR gyroscope OR inertial measurement unit OR IMU) AND ((angle OR rotation OR kinematic) AND (joint OR shoulder OR elbow OR wrist OR hip OR knee OR ankle OR foot))

This alone produced 1,468 results. The subsequent search process included removing duplicate works and reviewing titles and abstracts with the following initial inclusion and exclusion criteria:

- Must be a journal article
- Must be published between 2000-2018
- Data collection must include living human subjects
- Cannot require invasive method for alignment (e.g., bone screws or surgery)
- Method cannot be indirect (e.g., inferring joint kinematics from other sensors)
- Method for estimating IMU orientations cannot depend on external sensor information (e.g., Microsoft Kinect)

Then, the methods sections for each publication were reviewed with the following additional inclusion criteria:

- If estimating at least one joint angle, at least one accelerometer or angular rate gyro must be attached to each body segment *or* one body segment must be sufficiently stationary during testing
- Must include information about how the sensor frames were aligned to the body segment to which they were attached

Following these exclusion and inclusion criteria, 72 publications remained. A manual search of the included articles' references and citations lists was also conducted. The final number of publications found as a result of this systematic survey is 112.

4.2.3 Results and Discussion

Figure 4.1 illustrates the breakdown of publications that estimated rotations across the joints included in this survey, namely the shoulder, elbow, wrist, finger, hip, knee, and ankle. Pertinent to this dissertation, more than half of the publications (73) included estimations of knee rotations.

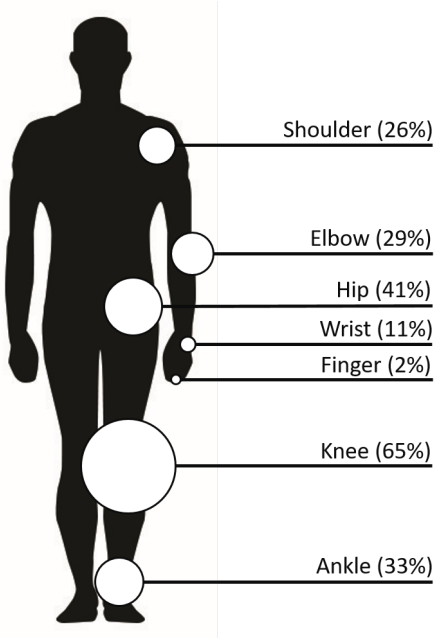


Figure 4.1: Breakdown of publications by joint.

The methods reported in the literature for defining anatomical frames of reference for inertial motion capture can be broadly categorized into one of four approaches:

1. Functional Alignment Movements (FAM): Participants complete some known movement(s) or pose(s) for which at least one anatomical axis can be estimated in an IMU's body-fixed frame of reference [70, 74, 80, 106, 115–162].

2. Assumed Sensor Orientation (ASO): An IMU is attached to a body segment such that the body-fixed frame of the sensor is approximately aligned with the anatomical frame of the body segment to which it is attached [52, 163–208].

3. Self-Aligning (SA): Body segment anatomical axes are estimated by exploiting a kinematic model or an assumption of the joint [19, 209–214].

4. External Information (EI): A source other than the IMUs (e.g., optical motion capture) provides information needed to determine the relationship between the IMU frames and the body segment anatomical frames [72, 215–219].

Figure 4.2 illustrates the breakdown of these approaches across all 112 papers found in this survey.

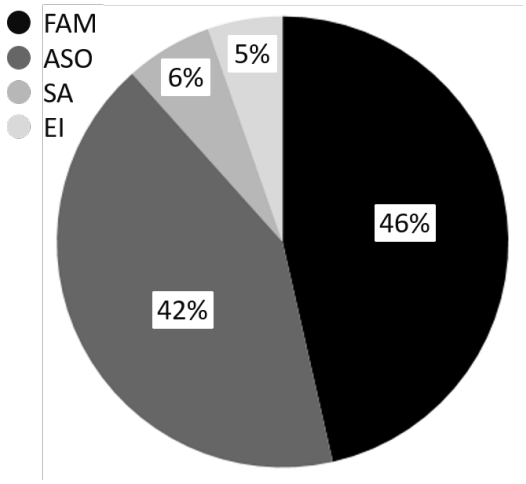


Figure 4.2: Breakdown of publications by approach, namely Functional Alignment Movements (FAM), Assuming Sensor Orientation (ASO), Self-Aligning (SA), and External Information (EI).

The great majority of the publications employed either FAM or ASO approaches, which is likely due to the fact that they are relatively simple to incorporate into an experimental protocol. While this breakdown is informative, Fig. 4.3 illustrates the frequency of these approaches over time.

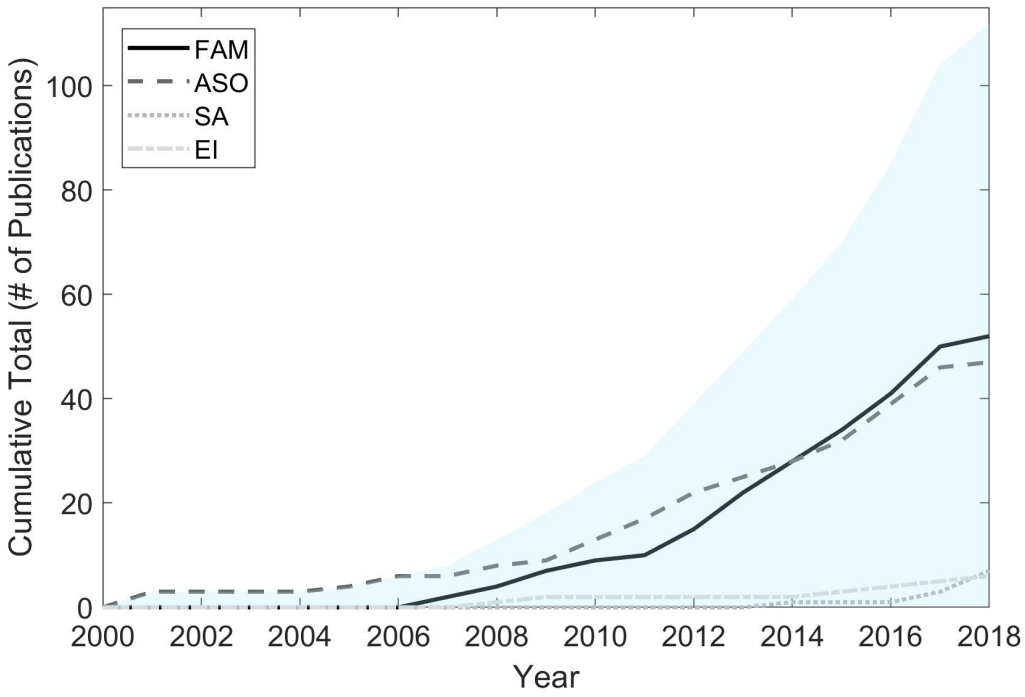


Figure 4.3: Cumulative total number of publications employing each method over time. The blue shaded area shows the cumulative sum across all approaches.

First, note the exponential growth in the number of studies either investigating new methods for estimating joint angles with IMUs or using those estimates to investigate differences between different subject populations. While the publications included in this systematic survey represent a very small slice of the studies employing inertial motion capture techniques, it is reasonable to assume that this growth also indicates the growing popularity of using IMUs for the biomechanics community as a whole. This trend further highlights the significant need to establish a convention for defining anatomical frames of reference for inertial motion capture.

Next, given the difficulties of estimating IMU orientation, it is reasonable that the earliest papers would use ASO methods followed by FAM methods. The SA methods have only just started to appear in the literature starting in 2014 with Seel et al. [19] whose work is the basis for defining the hinge axis of the knee discussed in this dissertation. Only a few researchers have used EI approaches for defining the frames, which is likely due to the impracticality of requiring additional motion capture modalities or equipment. Overall, it appears the community is still exploring different methods for defining anatomical frames of reference for inertial motion capture.

Focusing specifically on the knee, Fig. 4.4 illustrates which methods have been employed and by how many degrees of rotational freedom were estimated. The 44 publications in the '1D' category estimate solely flexion-extension rotations. The 5 publications in the '2D' category estimate flexion-extension and abduction-adduction rotations. Finally, the 24 publications in the '3D' category estimate flexion-extension, abduction-adduction, and internal-external rotations.

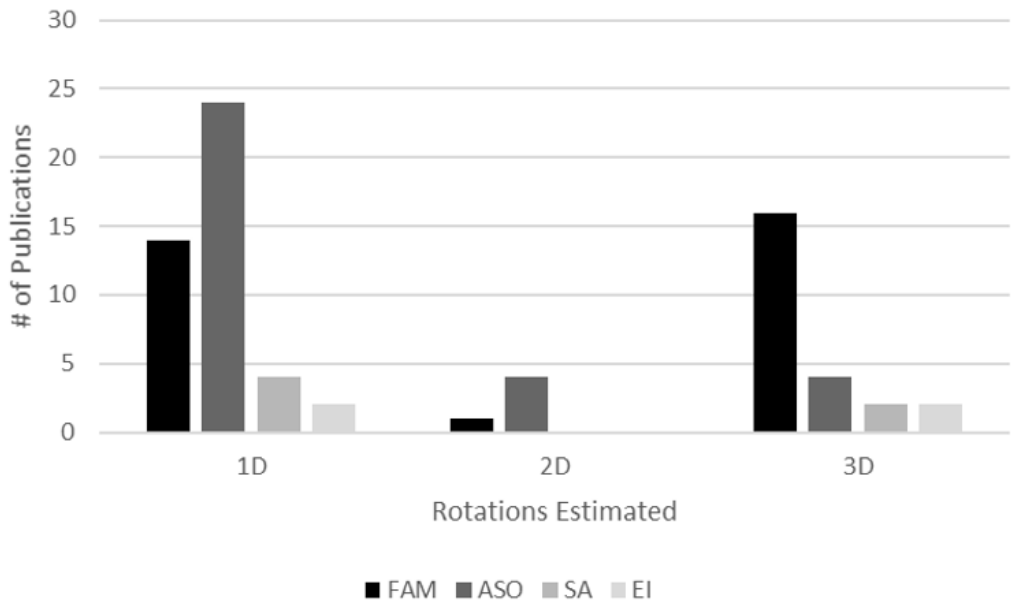


Figure 4.4: Breakdown of approaches by estimated rotational degrees of freedom across the knee.

From the 3D rotations results, FAM approaches are the most commonly used, though the exact movements employed vary across the studies. For example, Favre et al. [74] had participants stand in a neutral static pose followed by a hip abduction-adduction movement during which the knee is locked. These movements were used to define the superior-inferior and posterior-anterior axes, respectively, after which a subsequent cross product defines the complete orthonormal anatomical frame of reference. Fasel et al. [153] similarly had participants first stand in a neutral static pose to define the superior-inferior axes, then complete slow squats to determine the medial-lateral axes. Overall, these approaches tend to be consistent in estimating anatomical frames of reference, though there is frequently misalignment relative to frames defined using the (positional) optical motion capture conventions.

Only four publications employed ASO approaches to define anatomical frames of reference. Ahmadi et al. [194] reported validation results only for flexion-extension estimates, and Sun et al. [206] did not report any validation of their technique. The success reported by Kun et al. [175] was influenced by attaching the IMUs to rigid links strapped to the thigh and shank, which were connected by a universal joint. The accuracy demonstrated by Favre et al. [52] was dependent on expert placement of the IMUs on the body segments, which the authors described as being particularly difficult for the knee abduction-adduction (anterior-posterior) axis.

Finally, it is interesting that six of the seven publications using SA methods were specifically developed for the knee, one of which was proposed by Seel et al. [19] for a single rotational degree of freedom. While SA methods are appealing in that they do not require subjects to perfectly execute functional alignment movements, the results remain limited to date. Only two of the six SA publications for the knee aimed to estimate all three anatomical axes, and with the work by Bleser et al. [211] providing the foundation for Zimmermann, Taetz, and Bleser [214]. The most successful SA approach reported by Zimmermann et al. [214] implemented deep learning techniques trained on real and simulated IMU data to determine the anatomical frames. Although the results may hold promise, the methods would likely benefit considerably from training the algorithms using truth data collected with stereoradiography/dual-plane fluoroscopy techniques.

4.2.4 Conclusion

The systematic survey of the methods used to define anatomical frames of reference for inertial motion capture provides compelling evidence that there does not currently exist a common convention. In fact, there is evidence that not only has the biomechanics community as a whole not converged on a specific method, researchers are continuing to explore new approaches. In general, SA approaches are promising in that they do not depend on the researchers' ability to attach the IMUs to the body segments in any specific orientation (as required in ASO) or for the participants to properly execute functional calibration/alignment movements (as required in FAM). However, the SA approaches are relatively new and likely require significant validation (and training data in the case of [214]). Thus, the FAM approach is implemented in the next study to evaluate the measurement theory developed in Chapter 3 when applied to human subjects. However, it could also be readily extended to include future SA approaches.

4.3 Human Subject Testing

Given the successful results reported in Chapter 3, the remaining task for this chapter is to evaluate the efficacy of the measurement theory when applied to human subjects. There are two facets to this evaluation. The first considers the challenge of estimating the anatomical frames of reference using IMUs that are consistent with the definitions for optical motion capture. Following prior successes of employing functional alignment movements as revealed in the systematic survey, the participants in this study will also complete a set of movements to determine the anatomical frames. The optimization-based alignment technique from Chapter 3 provides a way of evaluating the success of this approach. The second facet considers the efficacy of the assumption that the knee predominately acts a hinge, or at least frequently enough for the measurement theory to provide reliable results. This assumption will be tested by estimating 3D joint angles during increasingly dynamic movements during which the hinge assumption would presumably begin to deteriorate.

In the past, the vast majority of studies conducted to validate 3D knee rotation estimations have done so using low dynamic, cyclical movements like walking or stair climbing. In fact, only two studies from the survey considered more dynamic movements. Fasel et al. [153] validated their algorithm on two minutes of alpine skiing on an indoor skiing carpet. They reported excellent results for flexion-extension and abduction-adduction and fairly good results for internal-external rotation (the anatomical angle most dependent on orientation about vertical). Al-Amri et al. [160] validated the Xsens MVN BIOMECH IMU-based system on walking, squatting, and vertical jumping with the last two exercises generally inducing larger ranges of motion. They reported excellent results for flexion-extension across all three tasks and "acceptable" results for the other two anatomical rotations. While the majority of applications of inertial motion capture focus on human gait analysis, there is an increasing need to investigate the differences between the two motion capture modalities (MOCAP versus IMU) during other highly dynamic movements. Specifically, it is likely that errors due to skin artefacts will manifest differently in the two motion capture modalities (see, for example, [220]).

4.3.1 Methods

Experimental Procedure

Testing with human participants (2M, 1F; age = 26.6 $\pm$ 1.9) is conducted over a range of increasingly dynamic tasks to expose differences in 3D rotations across the knee using both IMU and MOCAP measurement modalities. Participants' thighs and shanks (Fig. 4.5.) were outfitted with four IMUs sampling at 128 Hz (Opal sensors, APDM, Portland, OR, USA; sensor characteristics available at `http://www.apdm.com/wearable-sensors/`). Additionally, a 20-camera motion capture system (Motion Analysis, Santa Rosa, CA) sampling at 120 Hz tracked a MOCAP marker set made up of 67 markers [221] (visible in Fig. 4.5) to create a lower body model. This data set was collected for multiple purposes, one of which was for this validation study for IMU estimated 3D knee rotations. The University of Michigan Institutional Review Board approved the study and participants gave informed consent.

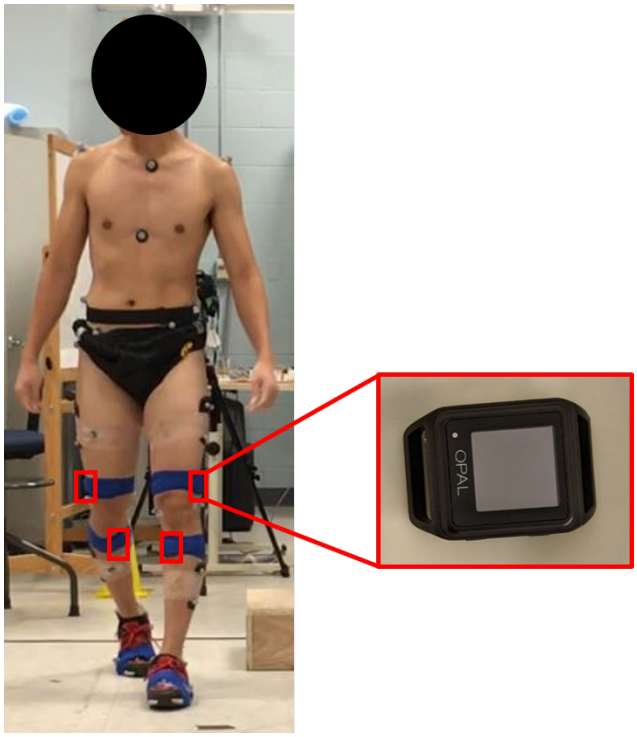


Figure 4.5: Participant outfitted with IMUs and MOCAP markers with a call-out of an IMU on the right.

Participants first completed three functional alignment movements to define the anatomical frames of reference for the thigh and shank. The first movement is a bicycle pedaling motion in which participants move their foot in an approximate elliptical trajectory in the sagittal plane. This movement is used to determine the medial-lateral axes of rotation using the procedure outlined by [19] and [20]. The second movement is a static period where participants stand in a neutral pose as still as possible. For the final movement, participants lock their knee and rotate their leg about the superior-inferior axis of the hip (and femur) while retaining the heel on the floor as a pivot point. Ideally, when the knee is locked the superior-inferior axis of the femur is aligned with the superior-inferior axis of the tibia. These latter two movements are used to determine the superior-inferior axes of the thigh and shank, which will be compared to each other.

The participants then completed a series of dynamic tasks including step ups, bicycle pedaling, drop landings, and jump cuts. This sequence of movements was selected for several reasons. First, the movements induce dominantly 2D motions (step ups, bicycle pedaling, drop landings) motions as well as motions with significant 3D components (jump cuts). Second, the movements span considerable dynamic ranges including slow to moderate steady-state motions (bicycle pedaling, step ups) to highly dynamic transient motions (drop landings, jump cuts). Thus, this sequence of motions will likely reveal possible challenges to the measurement theory. A summary of each movement follows.

A procedure similar to that described in [222] was used for the step-ups. The participants started from a neutral static pose with feet on the floor approximately shoulder-width apart, stepped up on a box (30 cm tall) with their left leg first before returning to the neutral static pose on top of the box. The participants then took a small step forward to the edge of the box, paused for a moment, then stepped down with their left leg first. They repeated this sequence five times and then completed the same protocol with the opposite leg leading the step up/down.

Prior to using the stationary bicycle, the proper saddle height was established using a goniometer. The saddle height was adjusted to achieve knee flexion between 27° and 37° in the starting position [223]. The participants were instructed to pedal while maintaining 70 RMP for one minute using the built-in bicycle speedometer for reference (CycleOps, 400 Pro Indoor Cycle).

For single leg drop landings, the participants began by standing on one leg (test leg) on a (30cm) step with their hands at their hips and with the contralateral leg slightly flexed at the knee. Participants then hopped off the step and landed on the test leg and were instructed to keep their eyes forward and their contralateral knee bent behind the test leg. The participants first observed this task demonstrated by the experimenter and then completed two practice trials prior to the test trials. A test trial was considered successful if the subject maintained balance on the test leg for three seconds on a force plate. Testing continued until ten successful trails were obtained, and then repeated using the opposite leg [224].

For the jump cuts, the participants jumped forward on a level surface starting

with both feet a distance 120% of their leg length onto a force plate, landed with their left leg and cut to their right. A total of five trials were collected, and then the procedure was repeated using the right leg for cutting. The criteria for a successful trial included: 1) jumping forward using both feet, 2) making full contact with the force plate, 3) absence of hopping on the force plate, and 4) cutting in the indicated direction [225].

MOCAP Calculations

The motion capture system (Motion Analysis, Santa Rosa, CA) included 20 cameras sampling at 120 Hz. After processing the marker position with a Butterworth filter (4^{th} order, $f_{cutoff} = 6$ Hz), a model was constructed in Visual3D (C-Motion, Germantown, MA) following ISB convention [113, 114]. Local coordinate systems and rotations across the knee were decomposed also using ISB convention [18]. In particular, the total (instantaneous) rotation across the knee was computed for comparison to that estimated using the IMU data per the following measurement theory extending from Chapter 3. As mentioned previously, while MOCAP represents the standard in the human biomechanics community, the estimates provided by this motion capture modality are not ground truth. Thus, the comparison between MOCAP and IMU estimates are discussed in terms of *differences* instead of in terms of *errors*.

Extension of IMU Measurement Theory to Human Subjects

The anatomical reference frames depicted in Fig. 4.6 are established from the functional alignment movements as follows. First, the angular velocity data collected during bicycle pedaling functional alignment movement is used to define the medial-lateral axes of the thigh and shank anatomical frames ($\hat{\boldsymbol{X}}_{AT}$ and $\hat{\boldsymbol{X}}_{AS}$, respectively) using the procedure outlined in [19]. The data collected during this movement is also used to calculate subject-specific means and variances of the measurement described in Eqn. 3.49 for Case 1b) when the knee is functioning as a hinge when the knee joint is rotating. Measurements calculated using data from the experiments are used to determine instances when the knee is reliably acting as a hinge via Mahalanobis distances (Eqn. 3.32) and χ^2 tests. As a reminder, a squared Mahalanobis distance larger than the χ^2 value for n degrees of freedom at a significance level $\alpha = 0.05$

indicates that the sample is an outlier of the distribution, in which case the knee is not considered to be acting as a hinge.

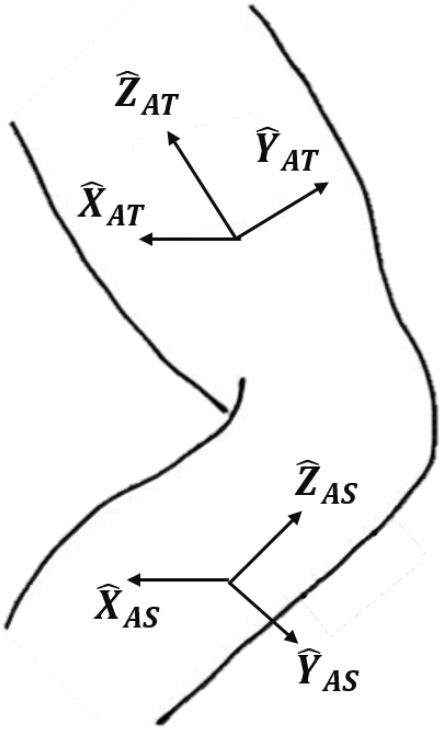


Figure 4.6: Anatomical frame definitions for the thigh and shank.

From the functional alignment data for the static neutral pose, the acceleration data is used to define the first estimate of the superior-inferior axes of the thigh and shank ($\hat{\boldsymbol{Z}}_{AT,1}$ and $\hat{\boldsymbol{Z}}_{AS,1}$, respectively). Specifically, the average direction of the measured acceleration during this still period is assumed to be aligned with the superior-inferior axes of the thigh and shank anatomical frames. The means and variances of the angular velocity and acceleration magnitudes are also calculated to define subject-specific static distributions to later determine whether the body segments are static or not via Mahalanobis distances and χ^2 tests. Finally, this acceleration data is also used to calculate subject-specific means and variances of

the measurement described in Eqn. 3.48 for Case 1a) when the knee is functioning as a hinge when the knee joint is stationary. Measurements calculated using data from the experiments are used to determine whether the knee is acting as a hinge via Mahalanobis distances and χ^2 tests.

Finally, the functional alignment movement where participants lock their knee and rotate their leg about the superior-inferior axis of the hip is used for secondary estimates of the superior-inferior axes of the shank and thigh based on angular velocity data instead of acceleration data. Specifically, a principal component analysis (PCA) [226] of the angular velocity returns as its first component secondary estimates of the superior-inferior axes of the thigh and shank ($\hat{\boldsymbol{Z}}_{AT,2}$ and $\hat{\boldsymbol{Z}}_{AS,2}$, respectively). Table 4.1 summarizes how both anatomical frames are defined.

Table 4.1: Definitions of the two anatomical frame definitions for the thigh. These descriptions are analogous to the shank anatomical frame definitions.

Frame	**Axis**	**Definition**	**Data**
1	$\hat{\boldsymbol{X}}_{AT}$	Knee Pedaling Axis	$\boldsymbol{\omega}_T$, $\boldsymbol{\omega}_S$
	$\hat{\boldsymbol{Z}}_{AT,1}$	Static Neutral Posture Axis	$\boldsymbol{a}_T$
	$\hat{\boldsymbol{Y}}_{AT,1}$	$\hat{\boldsymbol{Z}}_{AT,1} \times \hat{\boldsymbol{X}}_{AT}$	
2	$\hat{\boldsymbol{X}}_{AT}$	Knee Pedaling Axis	$\boldsymbol{\omega}_T$, $\boldsymbol{\omega}_S$
	$\hat{\boldsymbol{Z}}_{AT,2}$	Hip Internal-External Rotation Axis	$\boldsymbol{\omega}_T$
	$\hat{\boldsymbol{Y}}_{AT,2}$	$\hat{\boldsymbol{Z}}_{AT,2} \times \hat{\boldsymbol{X}}_{AT}$	

For both estimated anatomical frames, two cross products (preserving the medial-lateral axis) ensures orthonormal IMU-derived anatomical frames.

Data Analysis

The estimates provided by the ESKF from Chapter 3 yield the orientation of the shank IMU relative to the thigh IMU ($\boldsymbol{q}(t)_{T/S,IMU}$). Specifically, this quaternion describes the orientation of one IMU relative to the other with no regard to the IMU-based anatomical frames. By contrast, two different estimates of 3D knee

rotations are calculated using the two slightly different anatomical frames of reference summarized in Table 4.1. The first uses the anatomical frames defined with the static neutral pose in Eqn. 3.50 yielding $\boldsymbol{q}(t)_{AT/AS,IMU1}$. The second uses the anatomical frames defined with the hip internal-external rotation movement in Eqn. 3.50 yielding $\boldsymbol{q}(t)_{AT/AS,IMU2}$. Table 4.2 summarizes the different calculations using the anatomical frames defined in Table 4.1.

Table 4.2: Descriptions of the relative quaternion calculations.

	Relative Orientation	
Quaternion	**Of...**	**To...**
$\boldsymbol{q}(t)_{T/S,IMU}$	Shank IMU Frame	Thigh IMU Frame
$\boldsymbol{q}(t)_{AT/AS,IMU1}$	Shank Anatomical Frame 1	Thigh Anatomical Frame 1
$\boldsymbol{q}(t)_{AT/AS,IMU2}$	Shank Anatomical Frame 2	Thigh Anatomical Frame 2

The optimization technique described in Section 3.3.2 for determining body-fixed frame alignment is conducted three times. The first time uses $\boldsymbol{q}(t)_{T/S,IMU}$ to determine the "true" orientation of the IMUs relative to their respective body segments' anatomical frames defined by MOCAP. In other words, this result employs the MOCAP-determined anatomical frames. The second and third times use the two anatomical frames determined by the IMUs via the functional alignment movements (Table 4.2). Comparing these results allows one to determine misalignment between the IMU-derived anatomical frames and the MOCAP-derived anatomical frames. Determining these relationships provides direct comparison of the two IMU-based anatomical frames and aids interpreting the 3D anatomical rotations derived from the quaternions described previously. Table 4.3 describes the outputs from the optimization procedure for each of the relative quaternions in Table 4.2.

Finally, the previous experiments using the coordinate measurement machine only considered correlations and best fit lines between the IMU-derived estimates and those provided by the highly precise optical encoders. That analysis is appropriate given that the highly accurate encoders provide truth data. However, that

Table 4.3: Descriptions of the outputs from the optimization procedure relating the anatomical frames (AF) for each of the relative quaternions.

Quaternion	Optimization Output	Relative Orientation From...	Relative Orientation To...
$\boldsymbol{q}(t)_{T/S,IMU}$			
	$\boldsymbol{q}_{AT/T}$	Thigh IMU	MOCAP Thigh AF
	$\boldsymbol{q}_{AS/S}$	Shank IMU	MOCAP Shank AF
$\boldsymbol{q}(t)_{AT/AS,IMU1}$			
	$\boldsymbol{q}_{AT1/T}$	IMU Thigh AF 1	MOCAP Thigh AF
	$\boldsymbol{q}_{AS1/S}$	IMU Shank AF 1	MOCAP Shank AF
$\boldsymbol{q}(t)_{AT/AS,IMU2}$			
	$\boldsymbol{q}_{AT2/T}$	IMU Thigh AF 2	MOCAP Thigh AF
	$\boldsymbol{q}_{AS2/S}$	IMU Shank AF 2	MOCAP Shank AF

analysis is less appropriate in this study given that MOCAP cannot provide truth data for the reasons discussed in [107–111]. Therefore, Bland-Altman analyses are conducted to assess the agreement between the two motion capture modalities, each of which has error associated with their estimates [227]. However, decomposing the relative orientation of the anatomical thigh to anatomical shank into *Euler angles* (as described in prior chapters) will amplify the differences between IMU- and MOCAP-based anatomical frames. Therefore, a comparison of the *total rotation angle* (i.e., axis-angle representation [228]) is also included since it offers a comparison that is less sensitive to the differences between the anatomical frames for these two measurement modalities.

4.3.2 Results and Discussion

This section begins by examining the results for a representative trial for a single subject. Figure 4.7 illustrates all 4 quaternion components estimated for the thigh IMU (green) and the shank IMU (blue) for 16 minutes of data that span all four

movements (shaded regions) starting with the bicycle pedaling (approximately an hour into the data collection). These quaternion components represent the *absolute* orientation of each sensor relative to their respective world frames. Note that the low frequency oscillations present in the data derive from the orientation estimates drifting about the vertical axis, which is expected since the magnetometer data is not being used to correct the *absolute* orientation of the IMUs in the horizontal plane. The fact that the orientation estimates appear to drift at the same rate in the same direction is evidence of the measurement theory correcting the *relative* drift between the IMUs about the vertical axis, which is necessary to then accurately estimate the rotations across the knee as shown next.

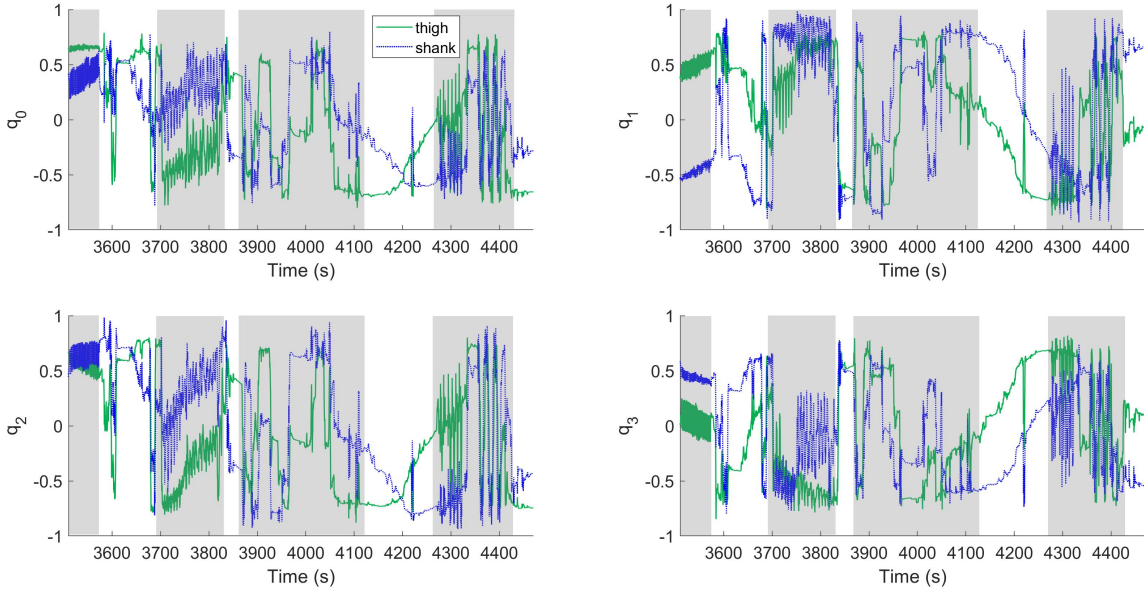


Figure 4.7: Quaternion estimates for the thigh (solid green) and shank (dashed blue). Note that the estimates from both sensors are drifting. The four shaded areas correspond to bicycle pedaling, box drops, jump cuts, and step-ups, respectively.

Next, Fig. 4.8 illustrates the resulting orientation of the shank anatomical frame relative to the thigh anatomical frame for the same sample trial by using the first anatomical frame described in Table 4.1. Doing so yields the IMU-estimated flexion-extension (FE), internal-external rotation (IE), and abduction-adduction (AA) of

the knee for this trial. Notice the absence of the low frequency drift error that was previously observed in Fig. 4.7. These sample results are significant in that they demonstrate the power of the measurement theory to estimate the *relative* orientation between sensors by correcting the *relative* integration drift error. While these results are for the illustrated 16 minute snapshot of data, the trend continues for the remainder of the trial. Now, the results comparing the IMU-based knee anatomical angles to those provided by MOCAP are presented.

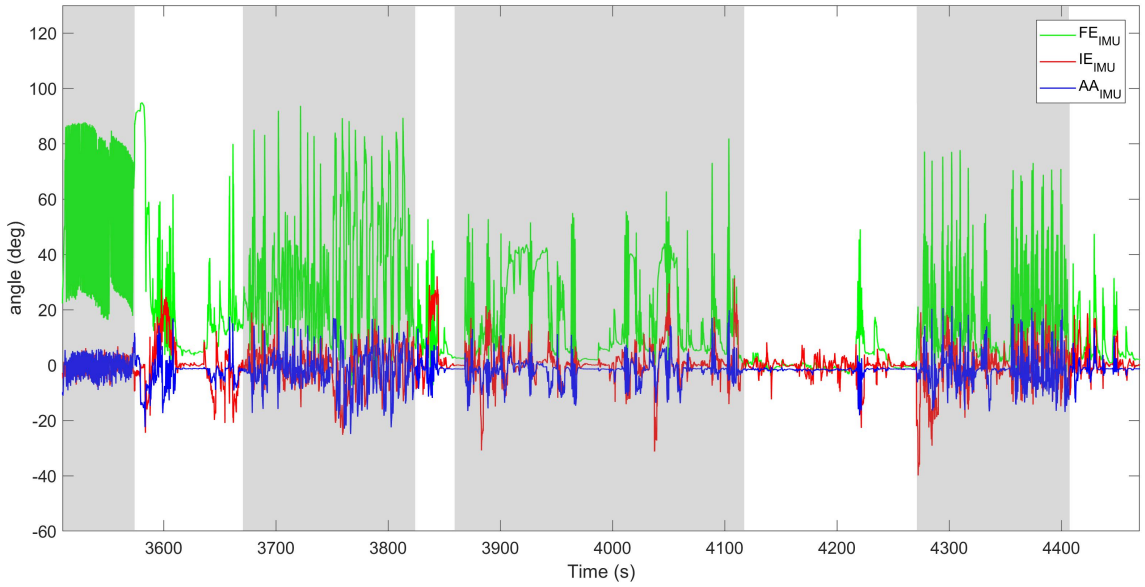


Figure 4.8: Estimates for flexion-extension (FE), internal-external rotation (IE), and abduction-adduction (AA) provided by the IMUs. The four shaded areas correspond to bicycle pedaling, box drops, jump cuts, and step-ups, respectively.

For the comparisons to MOCAP that follows, the IMU-based anatomical angles are derived from the optimization technique implemented on $\boldsymbol{q}(t)_{T/S,IMU}$ to determine the "true" orientation of the IMUs relative to their respective MOCAP-based anatomical frames. Figure 4.9 focuses on the quaternion estimates for the thigh and shank IMUs relative to their world frames for just the bicycle pedaling portion of the prior trial during which the subject pedalled at a constant speed for a minute. The bicycle pedaling should produce entirely periodic variations in the quaternions.

However, one clearly observes the superimposed linear variations due to integration drift error. While the integration drift error is clearly accumulating, it also appears to be accumulating at roughly the same rate for the thigh- and shank-mounted IMUs, which again is evidence of the measurement theory correcting the *relative* drift between the IMUs.

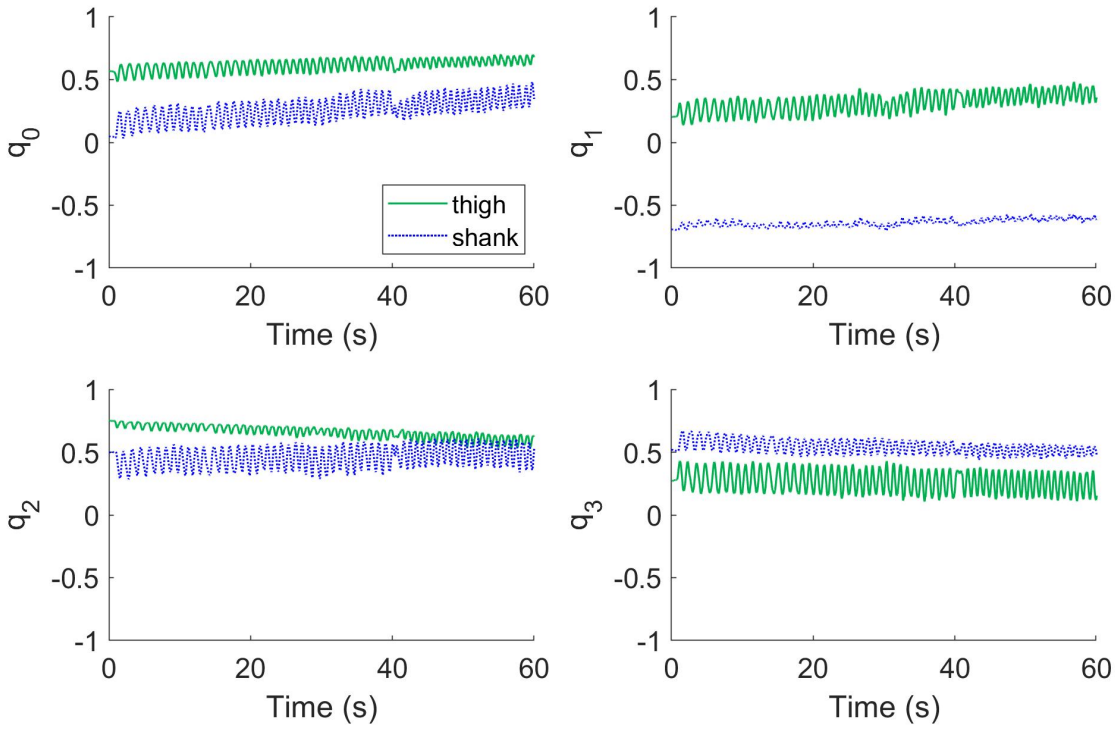


Figure 4.9: Quaternion estimates for the thigh (solid green) and shank (dashed blue). Note that the estimates from both sensors are clearly drifting.

Figure 4.10 illustrates the anatomical angles estimated by both IMU and MOCAP measurement modalities for the bicycle pedaling example above. Notice that the drift present in the results of Fig. 4.9 is now largely absent in the IMU-estimated anatomical angles of Fig. 4.10 due to the frequent drift correction over this time interval. In particular, the black squares and black triangles at the top of the figure denote time steps when the thigh and shank sensors, respectively, are considered to

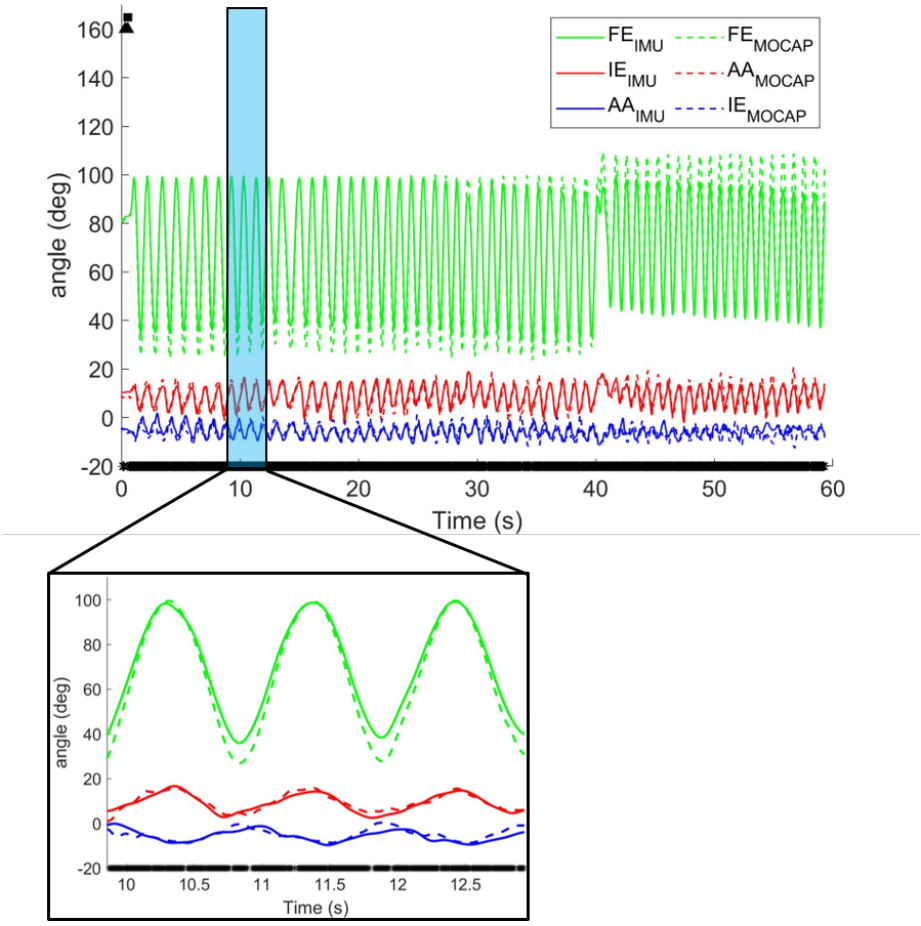


Figure 4.10: Estimates for flexion-extension (FE), internal-external rotation (IE), and abduction-adduction (AA) provided by IMUs (solid) and MOCAP (dashed). The black squares at the top denote times when the thigh is static, the black triangles at the top denote times when the shank is static, and the black asterisks at the bottom denote times when the knee is acting sufficiently as a hinge. The call-out at the bottom provides a magnified view of a small fraction of this time record where one can see the brief moments when the knee is not acting as a hinge (absence of asterisks). At around 40 seconds, the subject readjusted their posture on the saddle of the stationary bicycle, which is why the knee kinematics appear to change.

be static. However, these "static times" arise only at the very start of this time interval thus it is unsurprising that the flexion-extension (FE) estimates from the IMU

begin to drift near the end of this interval. Drift does not manifest in the internal-external rotation (IE) or abduction-adduction (AA) angles. By contrast, the black asterisks at the bottom of the figure denote times when the knee is acting sufficiently like a hinge to reliably correct the misalignment in the shank and thigh world frames. It is reasonable that the knee is acting as a hinge as frequently as is illustrated given that the medial-lateral axis and the distribution characterizing when the rotating knee is acting as a hinge is determined from the bicycle pedaling functional alignment movement. Additionally, since this anatomical kinematic constraint is a *vector* constraint, these updates provide drift error corrections for the IE and AA angles. Overall, the RMS differences between the IMU- and MOCAP-based estimates for FE, IE, and AA for this time interval are 5.1°, 2.4°, and 3.1°, respectively.

Completing this same analysis across all activities and subjects enables a full comparison of the anatomical angles by both measurement modalities in this study. Table 4.4 reports the resulting RMS differences in the anatomical angles as well as the total rotation angle for each activity. Interestingly, the differences between the IMU-based and MOCAP-based estimates were the greatest for the bicycle pedaling and jump cut activities. For the bicycle pedaling, these differences are largely driven by the FE estimates due to the aforementioned paucity of "static times" as in Fig. 4.10. For the jump cuts, the differences were also largely driven by the FE estimates as well. Nevertheless, the differences in the three anatomical angles and the total rotation angle across all activities remains below 5°.

Table 4.4: RMS angular differences between IMU and MOCAP estimates for each anatomical angle and total rotation angle for each activity type and across all trials and subjects.

Angle	**Step-Ups**	**Bicycle**	**Drops**	**Cuts**	**All**
FE	4.1	5.4	3.9	6.2	4.8
IE	4.3	2.4	4.6	3.5	3.7
AA	3.7	2.4	4.1	3.6	3.4
Total	4.4	5.5	4.1	6.1	4.9

An in-depth analysis of these differences follow from the results of the Bland-Altman analyses for each activity beginning with the step-ups. Figure 4.11 illustrates Bland-Altman plots for the three anatomical angles and the total rotation angle for the step-ups. The differences between the estimates provided by the two motion

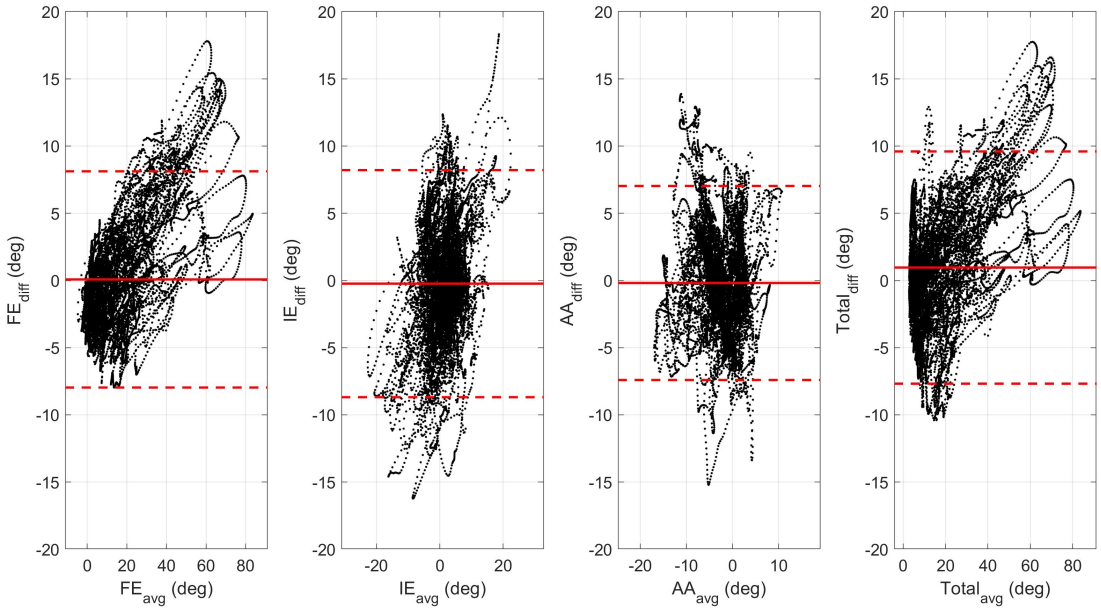


Figure 4.11: Bland-Altman plots for step-ups for flexion-extension (FE), internal-external rotation (IE), and abduction-adduction (AA) angles as well as the total rotation angle. The solid red line represents the mean difference and the dashed red lines represent the limits of agreement between the two motion capture modalities (1.96 times the standard deviation of the differences) [229].

capture modalities are plotted against their averaged estimates. For the anatomical angles, the mean differences are approximately zero. However, there is a slight offset in the total rotation angle indicating that, on average, the IMU method yields similar but consistently smaller estimates than MOCAP. The cause could be explained by the FE results for which there appears to be a proportional bias between the two modalities. This proportional bias is even more evident in the bicycle pedaling task.

Figure 4.12 illustrates the analogous Bland-Altman plots for the bicycle pedaling

task. Note that the average FE range of motion for all three subjects estimated by MOCAP is 88° whereas that estimated by the IMUs is 82°. While the difference between these predicted ranges of motion is not statistically significant given the range of agreement apparent in Fig. 4.12, the IMU method yields similar but consistently smaller FE estimates than MOCAP. However, this is not the case for the IE or AA angles. In fact, the IE and AA predictions have smaller ranges of agreement for the bicycle pedaling than for the step-ups.

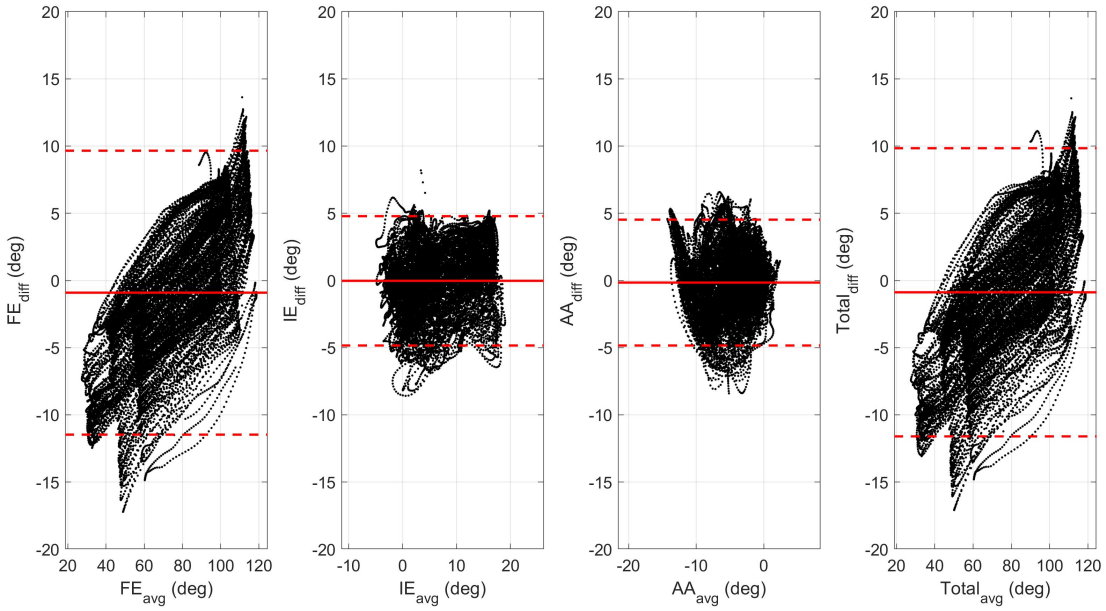


Figure 4.12: Bland-Altman plots for bicycle pedaling for flexion-extension (FE), internal-external rotation (IE), and abduction-adduction (AA) angles as well as the total rotation angle. The solid red line represents the mean difference and the dashed red lines represent the limits of agreement between the two motion capture modalities (1.96 times the standard deviation of the differences) [229].

Next, the Bland-Altman results for the box drops are illustrated in Fig. 4.13. As with the step-ups and bicycle pedaling, there is some evidence of the proportional error in the FE predictions. Also, the ranges of agreement for the IE and AA angles are of a similar magnitude as the step-ups.

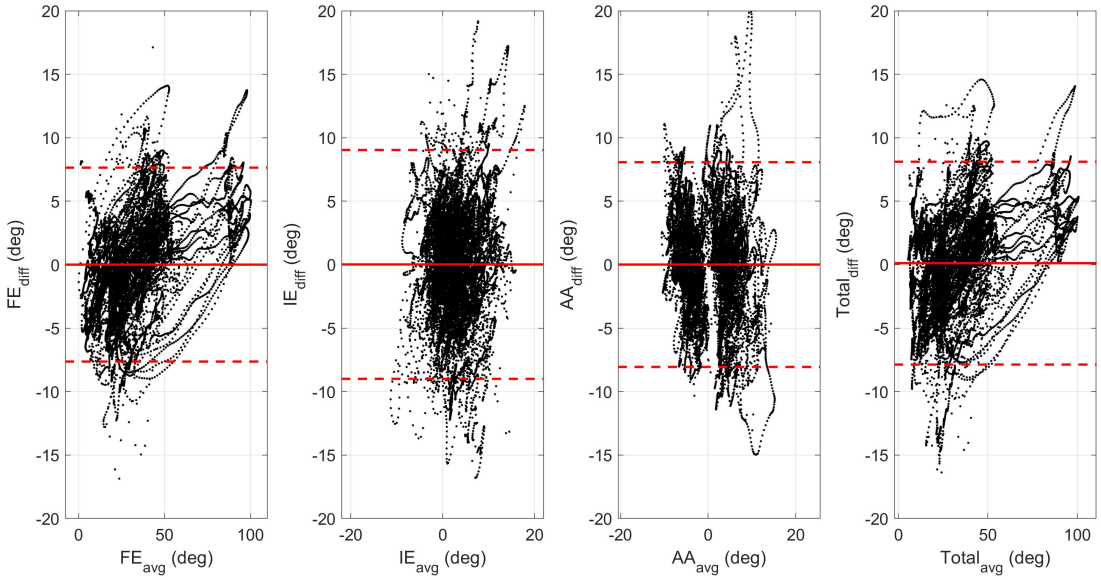


Figure 4.13: Bland-Altman plots for box drops for flexion-extension (FE), internal-external rotation (IE), and abduction-adduction (AA) angles as well as the total rotation angle. The solid red line represents the mean difference and the dashed red lines represent the limits of agreement between the two motion capture modalities (1.96 times the standard deviation of the differences) [229].

Lastly, the results for the jump cuts are similar in magnitude for the IE and AA angles whereas the proportional bias is again evident in the FE estimates. The FE estimates also drive the differences in the total rotation angles as well. Unsurprisingly, the results for the jump cuts exhibit the largest differences between the two measurement modalities due to the highly dynamic nature of this task. During this task, there are limited opportunities to correct the two IMU orientation estimates since the knee is neither static nor acting as a hinge,. In addition, this task would induce the greatest errors from skin artefacts.

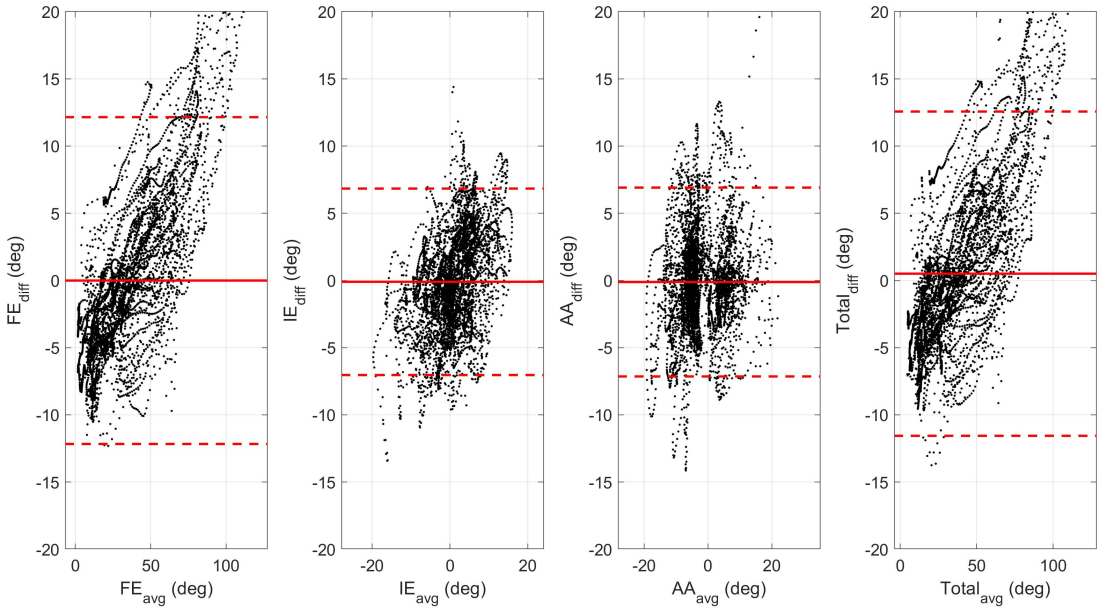


Figure 4.14: Bland-Altman plots for jump cuts for flexion-extension (FE), internal-external rotation (IE), and abduction-adduction (AA) angles as well as the total rotation angle. The solid red line represents the mean difference and the dashed red lines represent the limits of agreement between the two motion capture modalities (1.96 times the standard deviation of the differences) [229].

Finally, a brief discussion of the differences in the two IMU-derived anatomical frames (determined by the two sets of functional alignment movements) with that determined by MOCAP using the optimization technique of Chapter 3. Figure 4.15 shows the average angular difference between each of the IMU-derived anatomical frames (summarized in Table 4.1) and their respective MOCAP anatomical frames. These results are averaged across all trials, activities, and subjects. For the shank, there does not appear to be any significant difference between the two IMU-derived anatomical frames. Thus, using the acceleration measured during the static neutral pose (Anatomical Shank Frame 1) or the angular velocity measured during the hip internal-external rotation movement (Anatomical Shank Frame 2) yield essentially the same difference in the shank anatomical frames relative to that determined by

MOCAP (as measured by the total angle). For the thigh, however, the static neutral pose (Anatomical Frame 1) yields a results that more closely aligns with that of MOCAP.

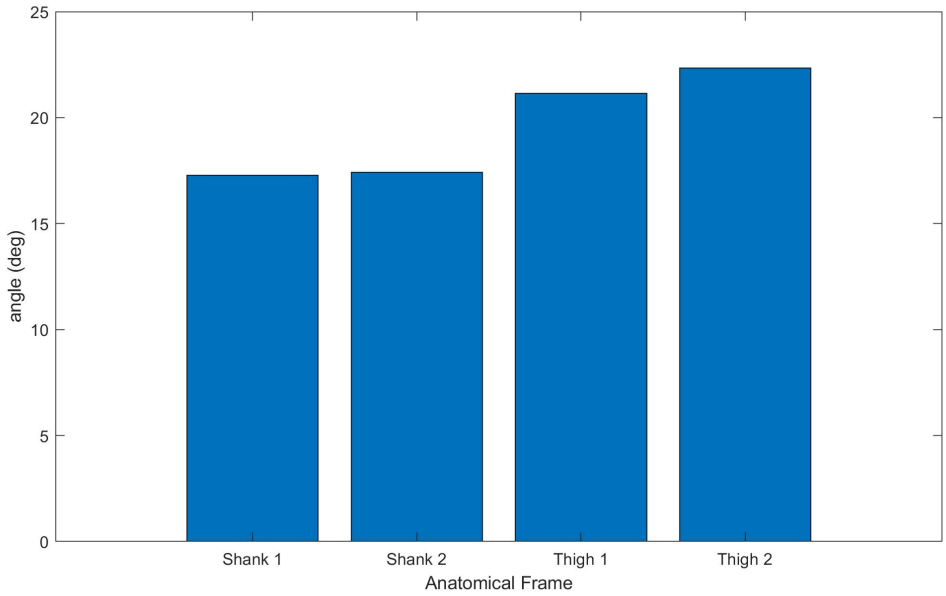


Figure 4.15: Total angular differences between the IMU-based anatomical frames and the MOCAP-based anatomical frame via the optimization technique of Chapter 3 averaged across all trials, activities, and subjects.

4.3.3 Conclusion

This study investigates the efficacy of the measurement theory to estimate three-dimensional rotations across the human knee. This is achieved by comparing the IMU-derived estimates for flexion-extension, internal-external rotation, and abduction-adduction with those estimated by convention optical motion capture (MOCAP) for three subjects across a range of four increasingly dynamic tasks, namely: 1) step-ups, 2) bicycle pedaling, 3) box drops, and 4) jump cuts. Recognizing that MOCAP should not be construed as providing ground truth results, the *differences* in the three knee angles obtained by these two measurement modalities are examined us-

ing Bland-Altman analyses. The following major conclusions are drawn from these analyses. The majority of the differences between the two measurement modalities derives from knee flexion-extension where the IMU-based estimates are the fact that the IMU-based estimates are typically smaller than those estimated by MOCAP (e.g., a 5° difference is typical). By contrast, smaller differences arise in the estimated knee internal-external rotation and abduction-adduction angles. This is particularly true for the bicycle pedaling and jump cut activities. Furthermore, the optimization technique reveals a consistent misalignment between the MOCAP-defined anatomical frames and the IMU-defined anatomical frames and regardless of which of the two sets of functional alignment movements are employed. This is not surprising given that the anatomical frames are defined using entirely different procedures. Thus, caution should be exercised in comparing (and even interpreting) the three anatomical angles estimated by these two measurement modalities. However, both methods should yield consistent results for the total angle across the knee as the total angle is independent of the definition of the anatomical frame, though it is likely that errors due to skin artefacts manifest differently in the two motion capture modalities.

4.4 Summary and Conclusions

The first major contribution of this chapter is a systematic survey of the prior methods used to define anatomical frames of reference using IMUs, and his leads to several important conclusions. Not only has the biomechanics community not converged to a common method, researchers continue to explore new approaches. In general, self-aligning approaches are promising in that they do not depend on the researchers' ability to attach IMUs to body segments in specific orientations (as required in the assumed sensor orientation method), or for the participants to properly execute functional calibration/alignment movements (as required in the functional alignment movement method). As the number of studies employing IMUs continues to grow exponentially, there remains a significant need to establish a convention for defining anatomical frames of reference for inertial motion capture.

The second major contribution is the extension of the measurement theory to hu-

man subject testing, and the subsequent comparison of IMU-estimated knee angles to those provided by conventional MOCAP. Regarding the measurement theory, the study results clearly demonstrate the success of the anatomical kinematic constraint to correct the *relative drift error* between the sensors over long periods of time. The major differences between the two motion capture modalities derives from the flexion-extension range of motion where the IMU method yields similar but consistently smaller estimates than MOCAP. This small difference might also derive from the optimization technique used to determine the orientation of the IMUs relative to the MOCAP-defined anatomical frames. More importantly, the results highlight the overall success of the measurement theory in aligning the two IMU world frames without any reliance on magnetometer data, which can often be untrustworthy. Finally, the study reveals that the two sets of functional alignment movements used to estimate the superior-inferior (internal-external rotation) axes of the thigh and shank lead to anatomical axes that are misaligned with those defined by MOCAP. This finding again underscores the need for the broader research community to develop a standard for determining anatomical axes using IMU-based techniques.

CHAPTER 5

Human Crawling Performance Metric Development and Evaluation

5.1 Introduction

As was noted in Chapter 1, the second objective for this part of the dissertation is to tap into the potential inherent to IMU technology by defining new (non-traditional) metrics for well-defined tasks that draw on the strengths of this technology. Thus, the studies in this chapter contribute a novel method for studying human crawling performance in a naturalistic environment both in terms of discriminating performance and evaluating the effects of additional body-borne load.

Crawling represents a form of quadrupedal gait often employed by humans, but insights can be gleaned from studying the crawling of reptiles, amphibians, and insects, among other organisms as well. A gait pattern, as defined by Song and Waldron [230], is "*the time and the location of the placing and lifting of each foot, coordinated with the motion of the body in its six degrees of freedom, in order to move the body from one place to another.*" Relative to this definition, McGhee [231] notes over 5,000 different possible quadruped gaits, including trotting, galloping, and crawling that are considerably more common than others. One common form of crawling gait common to many organisms consists of a diagonal interlimb pattern where the limbs (e.g., arms and legs) on opposite sides of the body swing in unison [232, 233]. Diagonal gait has further been shown to have implications in designing

controllers for quadruped robots, which may leverage gait transitions observed in nature to improve performance [234–236]. For example, several species of lizards exhibit diagonal gait when transitioning to high speeds, and it is further believed that diagonal gait affords greater stability and maneuverability [237].

In humans, diagonal crawling gait is commonly observed during infant development and marks important motor learning gains [232, 233, 238]. Infants generally transition from crawling with their abdomen contacting the floor to crawling with contact solely through their hands and knees, which introduces new neural, biomechanical, and task constraints contributing to the challenges of maintaining balance and support [239]. Despite these challenges, infants observed in a longitudinal study systematically converge to the same diagonal interlimb gait [239]. Crawling balance and dynamic stability follow from a base of support formed with as few as two limbs contacting the ground [240]. Diagonal crawling gait yields a diagonal base of support that remains nominally beneath the center of mass. Studies of infant crawling suggest that the diagonal gait pattern may also yield the most dynamically stable crawling gait as it minimizes lateral (e.g., side to side) variation of the mass center relative to the (diagonal) base of support [232]. Additional studies explore the effect of limb length on diagonal gait. For example, short-legged animals (e.g., amphibians and reptiles) adopt diagonal gait for greater stability per the aforementioned reason [241]. The same holds for infants who preferentially adopt a wide stance with respect to their shoulder width by placing their limbs further from the sagittal plane [233]. Relative to adults, infants have proportionally shorter limbs and adapt to relatively poorer balance by widening their base of support. When adults are forced to widen their base of support (limbs forced further from sagittal plane via obstructions), they converge to using the same diagonal crawling gait of infants [233].

Using this literature as a foundation, this chapter proposes metrics that define and discriminate human crawling performance (with particular attention paid to limb coordination) in the context of an outdoor obstacle course that embeds a crawling task for military personnel [16]. Crawling of military personnel may signal functional capacity following injury [21], expose the mechanisms and types of injury [22, 23], or be selected to optimize the size and configuration of carried loads [24–28]. Histori-

cally, crawling performance for military assessment is quantified solely by the time required to complete a crawling task. For example, the time to crawl a specified distance may serve as the measure of crawling performance. Although completion time is an important measure of performance, it does not reveal the techniques used or the underlying biomechanical motions that contribute to superior crawling performance. For example, studies [25, 26] conclude that large added loads from body armor significantly increases completion time. However, completion time does not reveal any of the companion changes in the biomechanical movements as also noted by observation [26].

This chapter expands prior studies of crawling performance by considering a broader range of performance metrics. In particular, crawling performance is studied across four measures of crawling performance; namely, crawl speed, crawl stride time, ipsilateral limb coordination, and contralateral limb coordination. This chapter begins by describing the development of the performance metrics, specifically detailing how they are derived from IMU data collected at the upper arms and thighs. The chapter then summaries two studies that: 1) investigate the ability of these metrics to discriminate superior crawling performance as summarized in [17] and 2) discriminate crawling performance as a function of body-borne loads as summarized in [242]. The consideration of crawling performance with body-borne loads is motivated by the need to understand the performance of warfighters with added load.

5.2 Methods

5.2.1 Participants, Equipment, and Protocol

For both studies, participants were recruited from a collegiate Reserve Officers Training Corps (ROTC) and club sports population (see Table 5.1 for demographic information) and completed numerous tasks embedded in a large outdoor obstacle course. One task, which required the participants to crawl swiftly on their elbows and knees, is often referred to as the "high crawl" in the context of military-style obstacle courses as described further below. All participants self-reported inexperi-

ence with the obstacle course and were therefore considered novices. The University of Michigan Institutional Review Board approved the study and participants gave informed consent.

Table 5.1: Demographics for recruited participants for both studies. Age, height, and mass are given as mean ± standard deviation

Study	**Number**	**Age (years)**	**Height (m)**	**Mass (kg)**
Study 1	33 (19M/14F)	20.2±2.0	1.75±0.12	71.6±4.2
Study 2	22 (15M/7F)	19.9±2.0	1.78±0.13	78.7±14.9

As mentioned above, participants wore an array of inertial measurement units (IMUs) (Opal, APDM, Portland, OR, USA). Four IMUs attached to upper arms and thighs with arbitrary orientation are pertinent to this study (Fig. 5.1).

Figure 5.1: Participant with four IMUs attached to upper arms and thighs (red boxes) and a callout of an IMU node on the right.

Each IMU includes a triaxial accelerometer (6g range, 14-bit resolution, 650 $\mu g/\sqrt{Hz}$ noise floor) and a triaxial angular rate gyro (2000°/s range, 16-bit reso-

lution, $0.03^{\circ}/\text{sec}/\sqrt{Hz}$ noise floor) sampled at 128Hz. Participants also carried a mock (plastic) rifle (mass 3.2kg, length 0.75m) with an IMU mounted to the barrel (Fig. 5.2(b)).

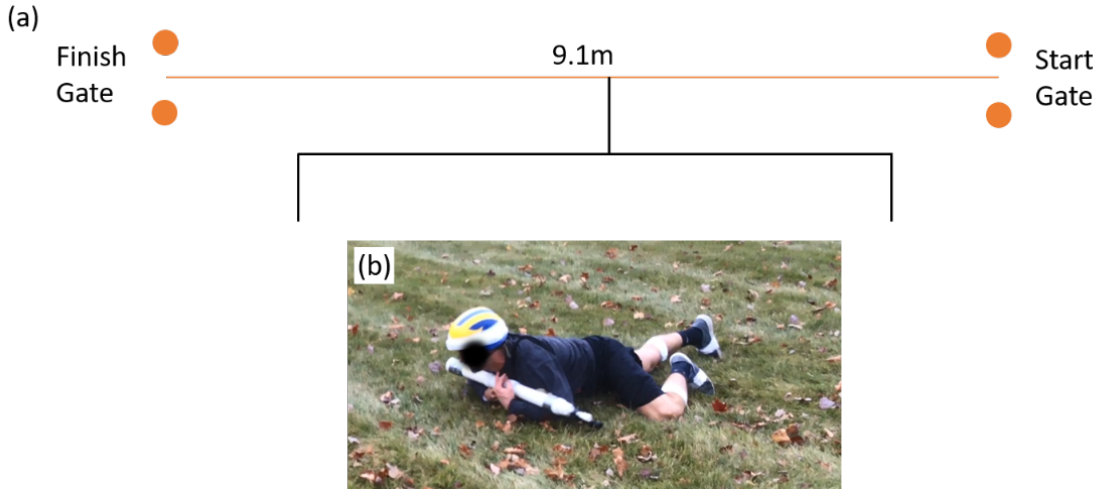


Figure 5.2: High crawl obstacle. **(a)** Layout of the high crawl obstacle, and **(b)** photograph of a participant completing the high crawl.

Participants completed a crawling task as part of a larger outdoor obstacle course modeled after the Load Effects Assessment Program (LEAP); see, for example, [16, 27]. The LEAP obstacle course is used by military organizations worldwide to evaluate the effects of load and other personal equipment on warfighter performance. It was created to replicate the common movements of warfighters and was specifically designed by the U.S. military to incorporate motions and tasks important for military personnel as emphasized in [16, 27]. In the context of the LEAP, the crawling task in this study is referred to as the high crawl. Participants started each crawl trial prone with their elbows even with a set of cones and were instructed to crawl as quickly as possible to another set of cones 9.1m (30ft) away while keeping their body as low to the ground as possible. In so doing, participants crawled on their elbows (not hands) and knees, though they also used their feet as additional points of contact when moving forward. Prior to testing, each participant walked through the course and received instruction on how to complete each obstacle. Si-

multaneously, they were encouraged to try each obstacle to familiarize themselves with the required movements and to reduce learning effects [27]. In the first study, participants completed the obstacle course once without additional load (Fig. 5.2).

The second study investigated the effects of added load on the crawl performance metrics developed from the first study. Thus, these participants completed the obstacle course multiple times in various loading conditions. For the loaded conditions, participants wore a (V-FORCE long) weight vest (Weightvest.com, Rexburg, ID, USA) laden with a select number of cast iron weights (mass 1.1kg, dimensions 10cm x 5cm x 4cm, equally distributed across front and back). Note that the participants in Fig. 5.1 and Fig. 5.2(b) both illustrate the unloaded condition (no load vest). Participants completed the crawling task a total of four times: two times without added load (unloaded condition: UL), once wearing an additional 15% of their body weight (loaded condition: 15BW), and once wearing an additional 30% of their body weight (loaded condition: 30BW). The added loads are considered relevant to understanding how body-borne loads influence the performance of warfighters. Participants performed over two testing sessions (on separate days) each of which included the UL condition and one of the two loaded conditions with 15 minutes in between for recovery. The order of loading conditions within and between testing sessions was randomized. In addition to the training period before testing, participants were encouraged to practice completing various obstacles (including the high crawl) during the first loaded condition to experience how the load may affect movement. For context, descriptive statistics for the additional loads are reported in Table 5.2. Note that the mean load representing 30% body weight (23.6 kg) is comparable to typical infantry fighting loads.

Table 5.2: Minimum, mean, maximum, and standard deviation (STD) of additional loads (in kg) secured to the participants.

Condition	**Min (kg)**	**Mean (kg)**	**Max (kg)**	**STD (kg)**
15BW	8.3	11.8	16.9	2.2
30BW	16.6	23.6	33.8	4.4

5.2.2 Analysis of IMU Data

Analysis of the IMU data begins by parsing the data into crawling gait cycles by detecting events that define the start and end of each cycle. Subsequent analysis of the parsed cycles reveals the speed of crawling and the coordination of the upper and lower limbs. These metrics, described next, build from prior measures for crawling (see, for example, [233, 237, 238, 243]). A description of each of these steps follows.

5.2.2.1 Crawl Gait Event Definition and Detection

The IMU data is parsed into individual gait cycles via the acceleration sampled at both upper arms. In particular, the acceleration magnitude at an upper arm exhibits a pronounced peak each time the associated elbow strikes the ground. These peaks readily define the start and end of each crawling gait cycle (Fig. 5.3). By contrast,

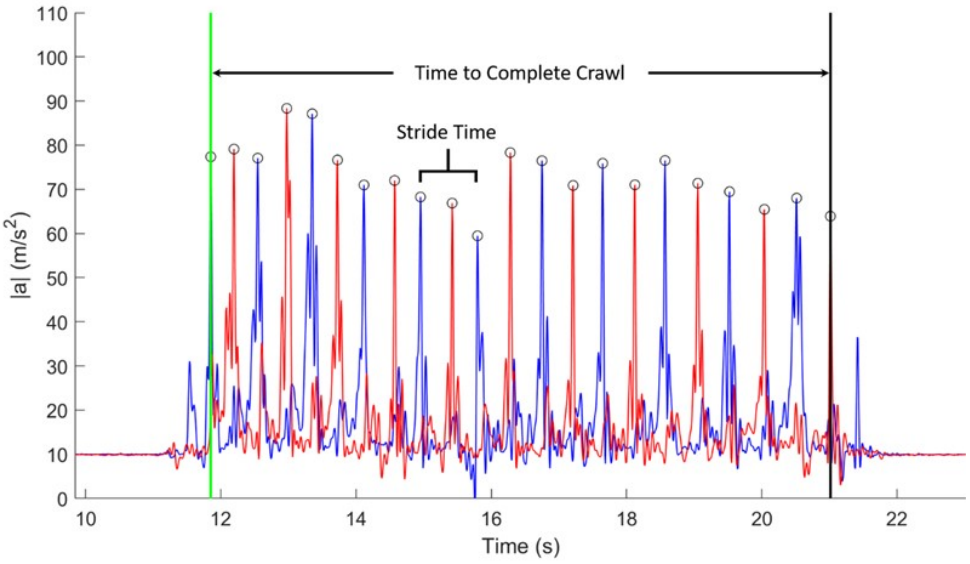


Figure 5.3: A plot of the acceleration magnitudes (including superimposed gravity) of IMUs attached to left (blue) and right (red) upper arm. Circles denote elbow strikes for each arm. The vertical green line indicates the start of the crawl and the vertical black line indicates the end of the crawl. A single "stride" time is illustrated for the left arm.

the acceleration sampled on each thigh does not readily distinguish gait cycles as the acceleration for the associated knee strikes can be significantly attenuated relative to the elbow strikes. Additionally, the knee often slips on the ground further confounding the detection of the cycle. Nevertheless, the motion of the thigh, and particularly the phasing of the motion of each thigh relative to both upper arms is critical to crawling performance. This phasing is captured by the coordination metrics proposed below.

5.2.2.2 Crawl Speed and Stride Time

The crawl task is bookmarked by the first and last elbow strikes, which define the crawl completion time (Fig. 5.3). The length of the crawl (9.1m) divided by the crawl completion time yields the average crawl speed. Stride times are defined from sequential elbow impacts on the same arm (i.e., time from one left elbow impact to the next left elbow impact); refer to Fig. 5.3. For each participant, the stride time is reported as the mean of the stride times for both right and left upper arms.

5.2.2.3 Ipsilateral and Contralateral Coordination Metrics

Two different (but related) metrics of crawl coordination are defined; namely ipsilateral and contralateral coordination. Ipsilateral coordination measures the phasing of limbs on the same side of the body (e.g., phasing of the right thigh relative to the right upper arm). Contralateral coordination measures the phasing of limbs on the opposite side of the body (e.g., phasing of the left thigh relative to the right upper arm). These coordination metrics are computed from the angular velocities measured synchronously on all four limbs as follows.

First, the principal axis of rotation for each upper arm during each stride is estimated from a principal component analysis (PCA) [226] of each upper arms angular velocity data. The PCA returns three orthogonal axes that describe the variation in angular velocity where the first axis captures the majority of the variation and is used as the principal axis of rotation. Additionally, the resulting principal axis nominally points in the lateral direction. The angular velocity for each upper arm

is projected onto the associated principal axis of rotation to yield "arm principal angular velocities" $\omega_{arm,i}$ where i=l or r for left or right upper arm. An analogous procedure is followed for the angular velocity measured by each thigh-mounted IMU, with one important difference. This PCA for each thigh is conducted twice, once using the strides defined by the left elbow strikes and then again using the strides defined by the right elbow strikes. Doing so yields "leg principal angular velocities" $\omega_{leg,ij}$ where i=l or r for left or right elbow strikes and j=l or r for left or right thigh. For example, $\omega_{leg,rr}$ is the right thigh principal angular velocities using the strides defined by the right elbow strikes and $\omega_{leg,lr}$ is the right thigh principal angular velocities using the strides defined by the left elbow strikes. In summary, each of these scalar quantities describes the angular speeds of each limb about the principal axis of rotation of each limb.

Next, a coordination matrix β (2 x 2) is formed that measures the phasing of each pair of arm and leg principal angular velocities. The components of the coordination matrix are

$$\beta_{ij} = \frac{\int_{t_1}^{t_2} \omega_{arm,i}(t)\omega_{leg,ij}(t)dt}{\sqrt{\int_{t_1}^{t_2} \omega_{arm,i}(t)^2 dt}\sqrt{\int_{t_1}^{t_2} \omega_{leg,ij}(t)^2 dt}} \tag{5.1}$$

where the limits of integration begin with the time of the first elbow strike (t_1) to the last elbow strike (t_2), i.e., the crawl completion time illustrated in Fig. 5.3. The component β_{ij} is a measure of the phase between the principal angular velocities of the i^{th} upper arm and the j^{th} thigh (calculated with the i^{th} elbow strikes), where again i, j = l(eft) or r(ight). Note that $-1 \leq \beta_{ij} \leq 1$ and that the limiting values $\beta_{ij} = 1$ and $\beta_{ij} = -1$ denote perfectly in-phase and perfectly out-of-phase motions, respectively; refer to illustrative example in Fig. 5.4 for principal angular velocities that are sinusoids.

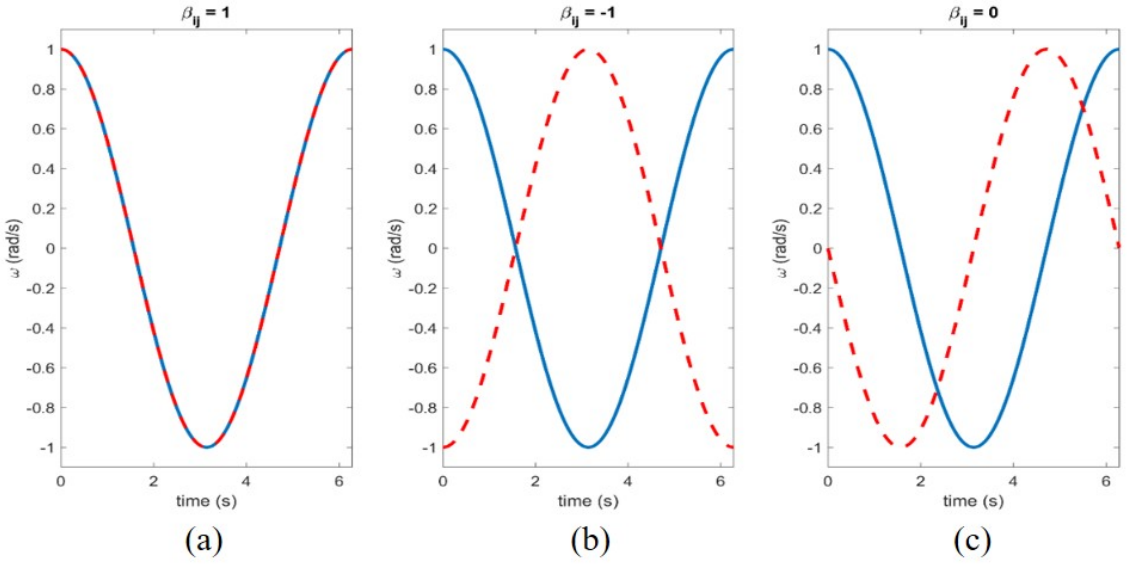


Figure 5.4: Example illustrating the interpretation of the components of the coordination matrix for principal angular velocities described by two (phase-shifted) sinusoids having period 2π. In this case, β_{ij} reduces to $cos\phi$ where ϕ is simply the phase angle between the two sinusoids. **(a)** β_{ij}=1 for two principal angular velocities that are exactly in-phase (ϕ=0). **(b)** β_{ij}=-1 for two principal angular velocities that are exactly out-of-phase (ϕ=π). **(c)** β_{ij}=0 for two principal angular velocities that are out-of-phase by ϕ=$\pi/2$.

However, the principal angular velocities during crawling are not simple sinusoids as illustrated in the example of Fig. 5.5. Shown are principal angular velocities for each of the four limb combinations over a typical stride for an exemplar high performer. The associated components β_{ij} of the coordination matrix are also reported, and they include values suggesting significant out-of-phase limb motions (-0.89, -0.72) as well as values suggesting significant in-phase limb motions (0.71, 0.59).

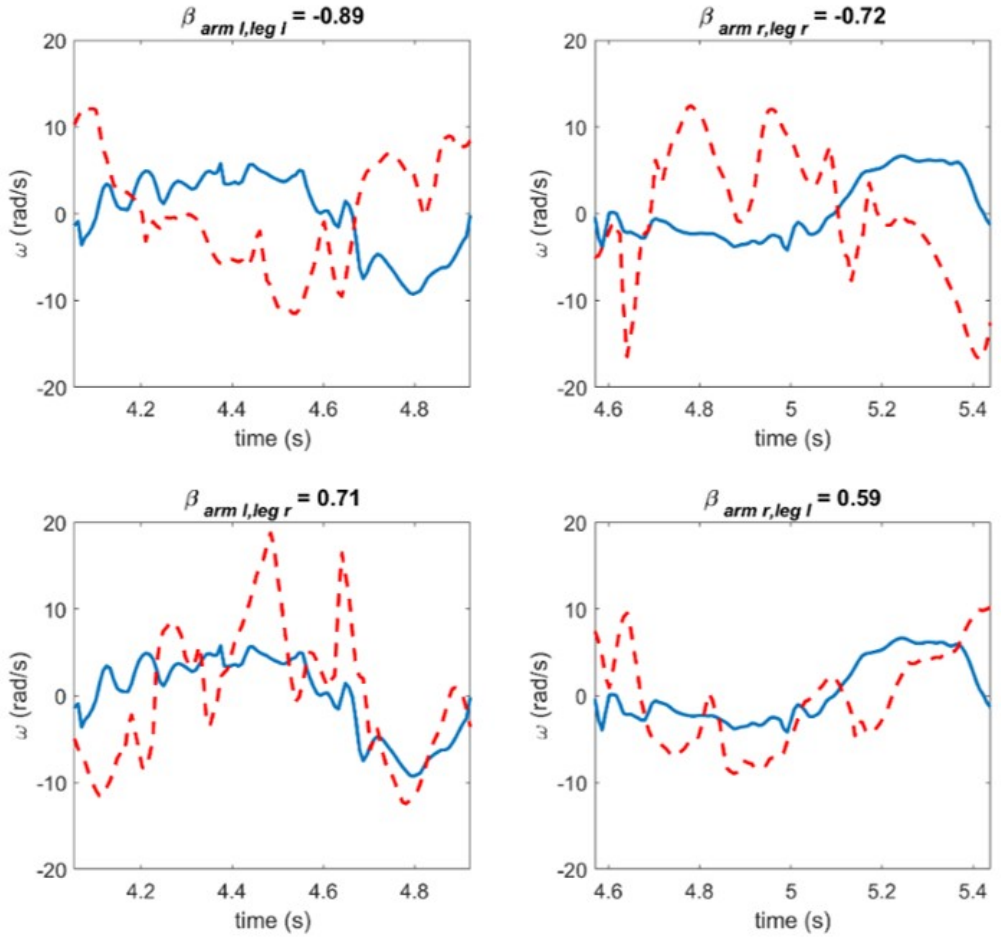


Figure 5.5: Representative principal angular velocities of the upper arms (blue) and thighs (dashed red) over a single stride for an exemplar high performer. If, for example, the right (left) upper arm is rotating counterclockwise (clockwise), the principal angular velocity is positive. The top row illustrates the relative phasing of the ipsilateral limbs and the bottom row illustrates the phasing of the contralateral limbs. The resulting components of the coordination matrix β_{ij} for this example stride are reported.

Reported next are the four components (β_{ij}) averaged across all strides and summarize those results in a "coordination map" as illustrated by the example shown in Fig. 5.6. The average values of β_{ij} (white circles) for each indicated pair of limbs are shown relative to the possible range of values $-1 \leq \beta_{ij} \leq 1$. In the coordination map,

positive (negative) values of β_{ij} that lie in the upper (lower) half of the range signify limb phasing that tends towards in-phase (out-of-phase). The coordination map of Fig. 5.6 is typical of a high performer as it reveals significant out-of-phase motion of the ipsilateral limbs ($\beta_{armr,legr}$=-0.79, $\beta_{arml,legl}$=-0.83) simultaneously with significant in-phase motion of the contralateral limbs ($\beta_{armr,legl}$=0.66, $\beta_{arml,legr}$=0.76).

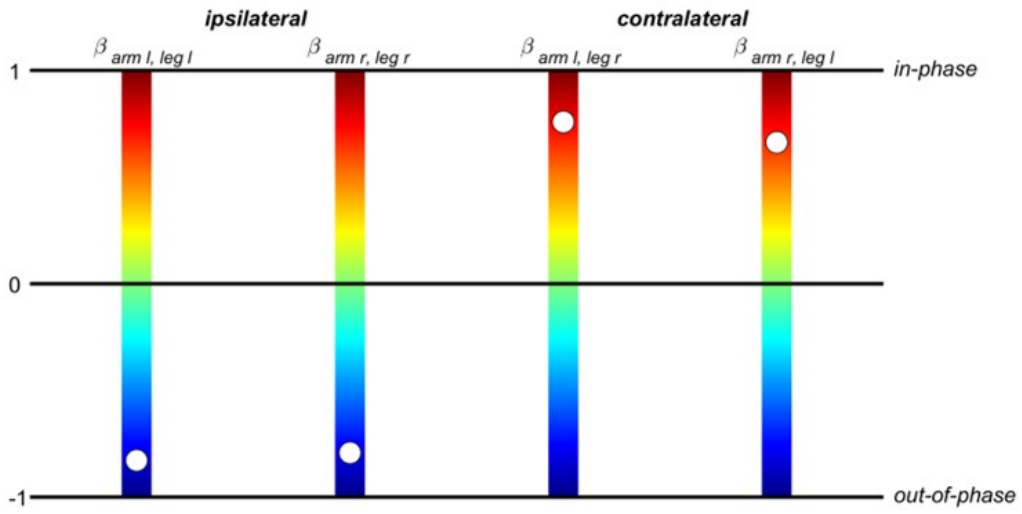


Figure 5.6: Coordination map for exemplar high performer from Fig. 5.5. Components β_{ij} (averaged across all strides) of the coordination matrix (white dots) are shown relative to possible range -1 (blue)$\leq \beta_{ij} \leq$1 (red). The upper half designates $\beta_{ij} >$0 for phasing that tends towards in-phase; the lower half designates $\beta_{ij} <$0 for phasing that tends towards out-of-phase.

Leveraging this observation, two overall crawl coordination metrics are defined that measure ipsilateral limb coordination

$$\beta_i = \frac{\beta_{armr,legr} + \beta_{arml,legl}}{2} \tag{5.2}$$

and contralateral limb coordination.

$$\beta_c = \frac{\beta_{armr,legl} + \beta_{armr,legl}}{2} \tag{5.3}$$

5.3 Study 1: Discriminating Crawling Performance

5.3.1 Cluster and Statistical Analysis

A k-means cluster analysis explored how crawl speed, stride time, contralateral coordination, and ipsilateral coordination contribute to overall crawling performance and technique. The optimal number of means is evaluated with a standard criteria known as the silhouette criterion [244], which provides a measure of how similar variables are within a cluster and how dissimilar they are across clusters. Prior to the cluster analysis, each of the four variables were normalized by their respective ranges to assign equal weighting in determining the clusters. After confirming normality, differences in the four variables between groups are further assessed with Welchs t-tests (to account for unequal variances between groups) with statistical significance evaluated at α=0.05. These statistical tests were conducted to confirm the distinction of the groups identified by the cluster analysis.

5.3.2 Results

The silhouette coefficient (0.63) indicated a reasonable structure [245] that distributed the participants into two groups ($n_1 = 26$, $n_2 = 7$). The ratio of inter-cluster distance to intra-cluster distance (calculated using Euclidean L2 norm distances [246]) for crawl speed, crawl stride time, ipsilateral coordination, and contralateral coordination are 1.18, 1.19, 1.23, and 1.25, respectively. A ratio greater than 1 indicates that the parameter significantly contributes to the group composition. Thus, in this case, each of the four variables contributes approximately equally. Table 5.3 below provides the descriptive statistics distinguishing the two clusters (or groups) from one another.

Table 5.3: Mean ± standard deviation of four crawl performance metrics for two groups identified by the cluster analysis. Performance metrics include average crawl speed, crawl stride time, ipsilateral coordination and contralateral coordination. The differences (Δ) denote Group 1 - Group 2. The t-statistic, p-value, and Cohens d effect size results from the Welchs t-test.

Metric	**Group 1**	**Group 2**	Δ	**t**	**p**	***d***
Crawl Speed (m/s)	0.7±0.2	0.5±0.2	0.2	2.80	0.02*	1.09
Crawl Stride Time (s)	1.3±0.2	1.8±0.6	-0.6	-2.59	0.04*	-1.30
Ipsilateral Coord.	-0.63±0.11	-0.27±0.20	-0.37	-4.72	$<0.01^{\dagger}$	-2.29
Contralateral Coord.	0.57±0.20	0.22±0.37	0.35	2.42	0.04*	1.18

Significant at $\alpha = {}^{*}0.05$, $^{\dagger}0.01$.

The first cluster (Group 1) exhibits faster crawl speeds (denoting superior crawling performance), shorter crawl stride times, and coordination metrics closer to ±1 (denoting more coordinated limbs) as compared to the second cluster (Group 2). Shapiro-Wilk normality tests and q-q plots for all variables in both groups did not indicate any significant deviations from the normal distribution. Table 5.3 also provides the results of the Welchs t-tests. Importantly, Table 5.3 provides the effect sizes (Cohen's d) for the differences across the groups as referred to in the following discussion. Differences in crawl speed, crawl stride time, and coordination across the groups reveal clear differences in performance. Moreover, the performance difference appears to manifest in differing crawling technique as identified by the coordination metrics discussed next.

5.3.3 Discussion

It was hypothesized that superior crawling performance (as defined by faster crawl speed) is associated with greater limb coordination and shorter stride time. Confirming this hypothesis, the cluster analysis revealed one cluster (Group 1) was characterized by faster crawl speeds, shorter crawl stride times, and greater limb coordination. Additional statistical tests were conducted to further confirm the hypothesis and the distinction of the groups identified by the cluster analysis. The

effect size for the crawl speeds (also average crawl completion time) suggests the two groups significantly differ by approximately one standard deviation (d=1.09). The other speed related metric, crawl stride time, also significantly differs between the two groups with the stride times of the high performers (1.3s) being significantly shorter than those of the low performers (1.8s). Prior literature suggests the most effective strategy to increase speed is to decrease the stance duration [238, 243], implying the duration of the swing phases largely remains the same across different speeds. If so, one can infer that high performers spend approximately half a second less (on average) in the stance phase compared to low performers. Note that the product of the average crawl speed and the average crawl stride time yields an estimate for the average crawl stride distance, which is 0.9m for both high and low performers. This lends further support to the belief that the mechanics of the swing phase alone would not distinguish high from low performers [238, 243].

The most prominent difference in performance is revealed by ipsilateral coordination. The substantial effect size (d=-2.29) suggests the mean ipsilateral coordination for the two groups differs by more than two standard deviations. Contralateral coordination also significantly differs between the two groups, but the associated effect size is about half that for ipsilateral coordination. To understand this from another perspective, Fig. 5.7 compares the coordination maps for an exemplar high performer (same as that of Fig. 5.6) adjacent to that of an exemplar low performer.

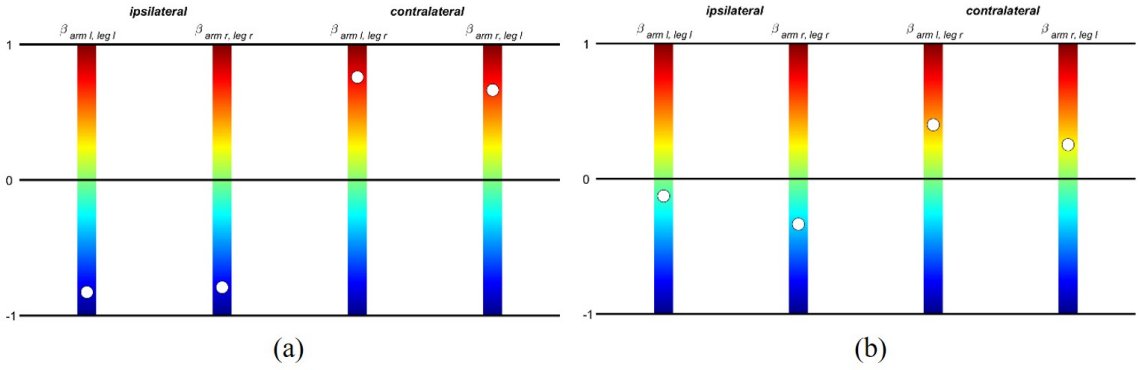


Figure 5.7: Example coordination maps for an exemplar **(a)** high performer (refer to Fig. 5.6) and **(b)** low performer.

For the exemplar high performer, the ipsilateral limbs are substantially out-of-phase (note the proximity of $\beta_{armr,legr}$ and $\beta_{arml,legl}$ to the blue end of the range where β=-1) while the contralateral limbs are substantially in-phase (note the proximity of $\beta_{armr,legl}$ and $\beta_{arml,legr}$ to the red end of the range where β=1). In contrast, for the exemplar low performer, the limbs are neither substantially in-phase nor out-of-phase (note the closer proximity of β_{ij} to the center of the map where β=0).

It is also interesting to examine possible cycle-to-cycle variations of these coordination metrics. To that end, Fig. 5.8 reports the moving averages (average of right and left first stride, right and left second stride, etc.) of the ipsilateral and contralateral coordination for (a) the exemplar high performer and (b) the exemplar low performer. The moving average for the high performer exhibits rather small

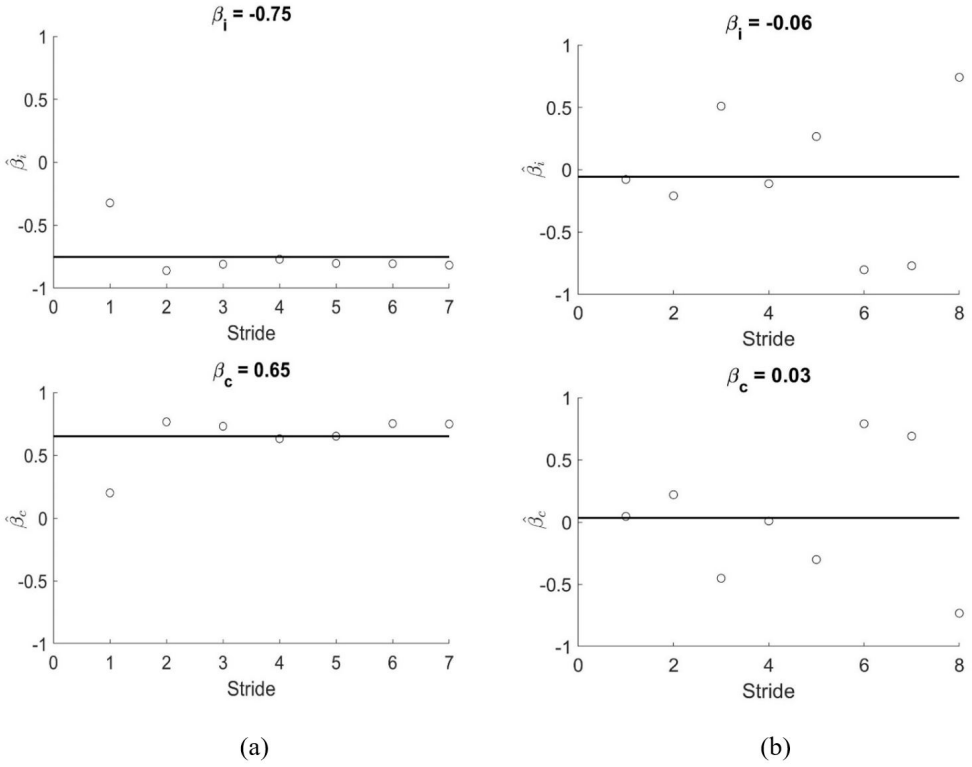


Figure 5.8: Moving averages of ipsilateral ($\hat{\beta}_i$) and contralateral coordination ($\hat{\beta}_c$) for the exemplar **(a)** high performer and **(b)** low performer. The black lines represent the averages across the trial as reported in each figures title.

variation. Thus, the high performer crawls with highly consistent coordination and also with coordination components reasonably close to ± 1 as illustrated by the overall average as well as in Fig. 5.7(a). Conversely, considerable variation is observed in the moving average of the coordination components for the low performer, which suggests that the low performer does not crawl with consistent coordination stride to stride. Moreover, the averages of the coordination components are close to 0, as illustrated by the overall average and in Fig. 5.7(b).

Overall, the exemplar low performer's limbs do not exhibit strong limb pairing and the limbs largely move independently of each another in cyclic sequence. Consequently, members in the low performing group incur significantly longer stride times as, for example, they wait to move their thigh only after the preceding elbow strike. In other words, low performers incur added stance time while they wait for their legs to catch up with their arms. In contrast, the exemplar high performers limbs remain strongly paired and with coordination that mimics the diagonal interlimb pattern often observed at faster locomotion speeds [233, 237]. Consequently, members in the high performing group crawl rapidly through superior ipsilateral and contralateral limb coordination. Indeed, the ipsilateral coordination is so strong that the ipsilateral limbs mimic the pairs of coupled oscillators previously observed in other modes of human locomotion (e.g., walking, creeping (i.e., crawling on hands and feet), and swimming) [247]. In all, the results from the additional statistical tests further confirm the hypothesis that higher performance is associated with more coordinated limbs and lower stride times.

One factor that may contribute to the ability to coordinate limbs could be the additional constraint of cradling the mock rifle in the arms. Cradling the mock rifle is an integral part of the testing protocol for the LEAP obstacle course (see, for example, [27]). In the exemplar high performer video (supplied in the Supplementary Material of [17]), the mock rifle remains firmly secured between the forearms and biceps, essentially creating a rigid constraint between the forearms. Perhaps as a result, the arm movements remain nearly perfectly out-of-phase. In the exemplar low performer video (as well as videos of other members of the low performance group) the mock rifle is secured using one hand and the crook of the contralateral

arms elbow, rendering it considerably less stable (i.e., not a rigid constraint). The added distraction and attention to stabilizing the mock rifle may contribute to the observed poorer limb coordination.

5.4 Study 2: Evaluating Load Effects on Crawling Performance

5.4.1 Statistical Analysis

Given the results from the first study, it was hypothesized that the performance metrics will be affected by load and that increased load results in degradations in performance as measured by those metrics. Since participants complete the obstacle course in all loading conditions (unloaded (UL), carrying 15% body weight (15BW), and carrying 30% body weight (30BW)), a repeated measures analysis of variance (ANOVA) framework was used to evaluate the effects of load on performance. Of the original 22 participants, 8 participants were eliminated from consideration based on the following exclusion criteria: 1) the participant failed to complete all loading conditions, 2) the data for one or more sensors was lost (sensor failure), or 3) the participant did not follow instructions. For each performance metric, the residuals were evaluated for normality with Shapiro-Wilk normality tests and q-q plots. The residuals were also checked for heteroscedasticity with Mauchly's test for sphericity. A violation of the sphericity assumption was handled by evaluating the F-statistic from the ANOVA with adjusted degrees of freedom via a Greenhouse-Geiser correction. Additionally, effect sizes (η^2) were calculated to quantify the magnitude of the effect that load has on the performance metric. Post hoc analyses via Tukey pairwise-comparisons were conducted for performance metrics with significant F-statistics. For each comparison, effect sizes (Cohen's d) were also calculated to evaluate the magnitudes of the differences between loading conditions. All statistical tests were evaluated at a significance level α=0.05. The relative magnitudes (as defined in [248]) of the effect sizes for the ANOVA and the Tukey post hoc analyses are summarized in Table 5.4.

Table 5.4: Relative magnitudes for the effect sizes for the ANOVA (η^2) and Tukey (d) analyses.

Effect Size	**ANOVA** η^2	**Tukey** d
Small	0.01	0.3
Medium	0.06	0.5
Large	0.14	0.7

5.4.2 Results

The results from the ANOVA reveal significant relationships with load for all four crawl performance metrics (Table 5.5). In particular, the effect sizes (η^2) indicate that the relationship between load and the performance metrics produces a large effect, the nature of which is described next.

Table 5.5: High crawl statistical results from the repeated-measures ANOVA (F-statistics with degrees of freedom, p-values, and effect sizes, η^2).

	F-statistic	**p**	η^2
Crawl Speed (m/s)	F(1.4,18.5)=43.2	$< 0.001^{\ddagger}$	0.77
Crawl Stride Time (s)	F(2,26)=18.1	$< 0.001^{\ddagger}$	0.58
Ipsilateral Coordination	F(2,26)=9.44	$< 0.001^{\ddagger}$	0.42
Contralateral Coordination	F(2,26)=4.25	0.03^{*}	0.25

Significant at $\alpha =$ *0.05, †0.01, and ‡0.001.

The results from the post hoc Tukey analyses are reported in Table 5.6. These pairwise comparisons reveal both the direction and magnitude of the differences between the loading conditions as discussed next.

Table 5.6: High crawl statistical results from all three post hoc Tukey pairwise comparisons (p-values and effect sizes, Cohen's d).

	UL-15BW		**UL-30BW**		**15BW-30BW**	
	p	***d***	**p**	***d***	**p**	***d***
Crawl Speed (m/s)	$< 0.01^{\dagger}$	0.55	$< 0.001^{\ddagger}$	1.12	$< 0.01^{\dagger}$	0.57
Crawl Stride Time (s)	$< 0.01^{\dagger}$	-0.68	$< 0.01^{\dagger}$	-0.98	0.03^{*}	-0.39
Ipsilateral Coordination	0.31	-0.32	0.07	-0.67	0.23	-0.33
Contralateral Coordination	0.27	0.28	$< 0.01^{\dagger}$	0.81	0.04^{*}	0.49

Significant at $\alpha = {}^{*}0.05$, $^{\dagger}0.01$, and $^{\ddagger}0.001$.

5.4.3 Discussion

The results from the Tukey post hoc analyses are further illustrated in the box plots in Fig. 5.9. The three pairwise-comparisons for the crawl speed are significant, implying the added loads increase the difficulty of the crawling task manifesting in significantly slower crawl speeds (Fig. 5.9a). This is consistent with previously reported results [24–28]. The three pairwise-comparisons for the crawl stride time are also significant (Fig. 5.9b), though the effect sizes indicate a potentially nonlinear increase in stride time with load. In particular, the effect size for the UL-15BW comparison (p<0.01, d=-0.68) is nearly twice that of the 15BW-30BW comparison (p=0.03, d=-0.39).

From the results from the first study, superior crawling performance is characterized by a diagonal stride pattern where elbows and knees on opposite sides of the body are in simultaneous ground contact [17]. This strategy results in ipsilateral limbs moving largely out-of-phase and contralateral limbs moving largely in-phase to produce faster crawl speeds and shorter crawl stride times [17]. Even though the ANOVA results for the ipsilateral coordination indicate a relationship with load, the Tukey analysis did not reveal any significant differences (Fig. 5.9c). This is likely due to variation in crawling technique meaning changes in this coordination metric may not be consistent across participants. The medium-large effect size for the UL-30BW comparison (p=0.07, d=-0.67) for ipsilateral coordination suggests load

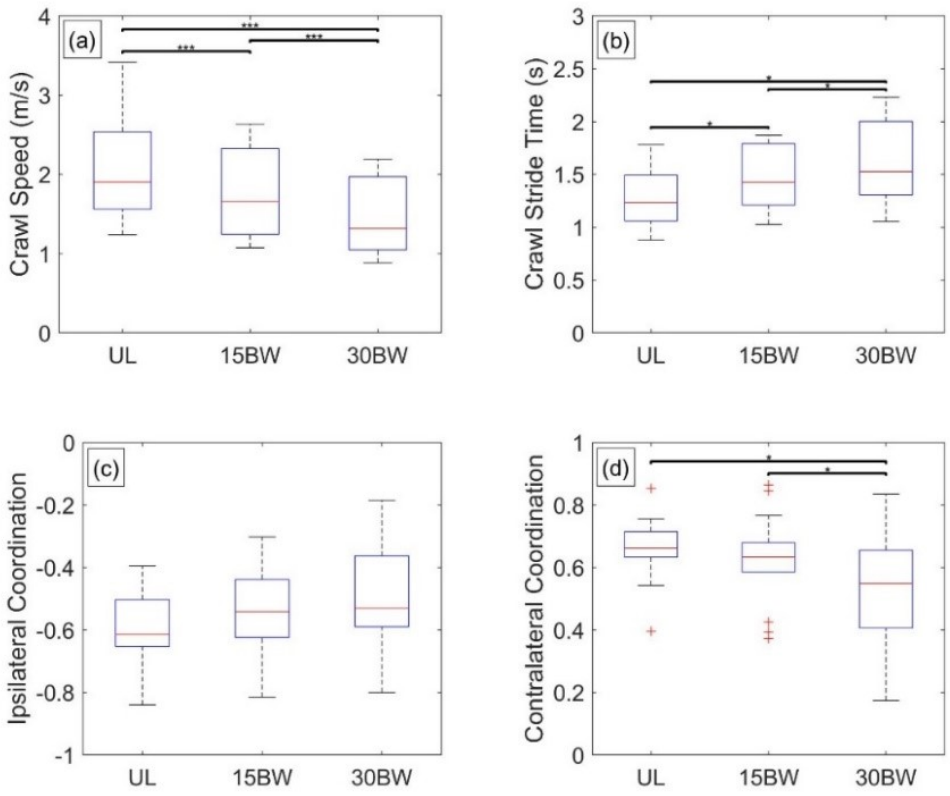


Figure 5.9: Boxplots depicting the results from the Tukey post hoc analysis for **(a)** crawl speed, **(b)** crawl stride time, **(c)** ipsilateral coordination, and **(d)** contralateral coordination. The bars denote significant differences between loading conditions at a significance level at α = *0.05, †0.01, and ‡0.001.

overall tends to reduce the degree to which the ipsilateral limbs move out-of-phase. Contralateral coordination, however, has significant relationships for the UL-30BW comparison as well as the 15BW-30BW comparison, which implies a possible nonlinear effect between load and the coordination of the contralateral limbs (Fig. 5.9d). In particular, the effect size for the 15BW-30BW comparison (p=0.04, d=0.49) is nearly twice that of the UL-15BW comparison (p=0.27, d=0.28). This could mean the 30BW condition is large enough that participants are unwilling and/or unable

to support the added weight using only two points of contact and instead maximize the time they have with three points of contact. This change in strategy is marked by a simultaneously decrease in contralateral coordination and increase in ipsilateral coordination yielding longer periods of three-point contact [233].

5.5 Conclusions

The first of the two studies contributes a novel method to measure and quantify crawling motion using four body-worn inertial measurement units attached to the upper arms and thighs. The data harvested from the embedded inertial sensors (accelerometers and angular rate gyros) were used to construct four measures of crawling performance; namely, crawl speed, crawl stride time, ipsilateral limb coordination, and contralateral limb coordination. Ipsilateral and contralateral limb coordination quantify the phasing of the rotations of the upper arms and thighs, and thus they are also a product of the underlying motor control (not directly measured). Collectively, these four metrics reveal the purposeful, coordinated movements that contribute to superior (or inferior) crawling performance as further revealed in a cluster analysis that aggregated participants into (statistically significantly different) high and low performance groups. The results of the cluster analysis and statistical tests confirmed the hypothesis that higher performance is associated with limb coordination and stride times. Overall, high performers exhibit superior limb coordination associated with a "diagonal gait" that produces faster crawl speeds and shorter crawl stride times. In contrast, low performers crawl at slower speeds with longer crawl stride times and with significantly less limb coordination.

The second of the two studies further demonstrates the ability of these IMU-derived performance metrics to quantify and characterize degradations in human crawling performance as the result of additional body-borne loads. In particular, the results revealed that added load decreases contralateral limb coordination thereby increasing crawl stride time and decreasing crawl speed. Though it was outside the scope of the study, one could readily study the effects of additional loading conditions (both in terms of magnitude and configuration) on crawling performance.

CHAPTER 6

Part I: Summary, Contributions, and Conclusions

Summary and Contributions

Part I of this dissertation advances the use of IMUs in the field of human biomechanics along two lines of research. The first contributes a new measurement theory for employing a pair of shank- and thigh-mounted IMUs to estimate the three-dimensional (3D) rotations across the human knee. The second contributes a method for deploying leg- and arm-mounted IMUs to understand human crawling performance. Summaries from both lines of research are detailed below.

Central to the measurement theory for the human knee is the following equation (Eqn. 2.9) for the rotation matrix defining the orientation of the shank anatomical frame relative to the thigh anatomical frame.

$$\boldsymbol{R}(t)_{AT/AS} = \boldsymbol{R}_{AT/T}\boldsymbol{R}(t)_{T/WT}\boldsymbol{C}(t)_{WT/WS}\boldsymbol{R}(t)_{WS/S}\boldsymbol{R}_{S/AS}$$

This rotation matrix, which yields estimates of the knee flexion-extension, internal-external rotation, and abduction-adduction angles, is decomposed into five component rotation matrices. Chapters 2–4 formulate each rotation matrix, starting in the middle and working outwards until the measurement theory is complete.

Chapter 2 contributes the foundational idea that the misalignment between independent world frames of the sensors, described by the rotation matrix $\boldsymbol{C}(t)_{WT/WS}$, can be estimated by exploiting the fact that the human knee can reliably be simpli-

fied as a hinge when certain criteria are met. When this is the case, the medial-lateral hinge axes defined in each sensor's (independent) world frame must be collinear. This condition is formulated as a vector kinematic constraint for aligning the sensor world frames and thus correcting for the inevitable relative drift error between these two frames. In Chapter 2, this kinematic constraint is treated in a deterministic manner. Results from the deterministic measurement theory are compared to ground truth measurements provided by optical encoders embedded in a coordinate measurement machine (CMM) that yield precise measurements of 3D rotations. Doing so yields estimates for the three rotation angles to within 4° RMS error.

The measurement theory is then extended in Chapter 3 to a probabilistic theory in anticipation of the challenges in using the deterministic theory for human subject testing. This begins by estimating the orientation of a single IMU relative to its world frame via a novel Error-State Kalman Filter (ESKF). After demonstrating the success of ESKF formulation for a single IMU (RMS errors of less than 5° relative to ground truth using the CMM), the ESKF formulation is then extended to two IMUs to estimate the rotation matrices $\boldsymbol{R}(t)_{T/WT}$ and $\boldsymbol{R}(t)_{WS/S}$ shown above. Importantly, replacing the deterministic treatment of the anatomical kinematic constraint from Chapter 2 with a probabilistic one yields a dramatic improvement in performance (RMS errors of less than 6° compared to RMS errors of more than 11°).

Lastly, Chapter 4 extends the measurement theory to human subject testing. Prior to doing so, the chapter opens with a systematic survey ($n = 112$) of the previous methods for defining anatomical frames ($\boldsymbol{R}_{AT/T}$ and $\boldsymbol{R}_{S/AS}$) using IMUs. The methods are classified as: 1) functional alignment movements, 2) assumed sensor orientation, 3) self-aligning, and 4) external information. In general, self-aligning methods are promising as they do not depend on the researchers' ability to precisely orient IMUs to the body segments (as required in assumed sensor orientation methods) or for the participants to properly execute functional calibration/alignment movements (as required in functional alignment movements methods). However, the self-aligning methods are relatively new and will require additional validation prior to adoption. Consequently, the more well-established functional alignment movements methods are employed in this study to determine the IMU-based anatomical

frames $\boldsymbol{R}_{AT/T}$ and $\boldsymbol{R}_{S/AS}$ above. The study then extends the probabilistic measurement theory developed of Chapter 3 to human subject testing for a range of increasingly dynamic tasks, namely: 1) step-ups, 2) bicycle pedaling, 3) box drops, and 4) jump cuts. Traditional MOCAP is simultaneously employed to provide independent, though not ground truth, estimates of the 3D knee angles. Overall, the measurement theory yields comparable 3D knee rotations to those estimated by MOCAP (RMS differences of less than 5° across all tasks and subjects). Central to this success is the measurement theory's ability to correct the *relative drift error* between the sensor world frames by exploiting the anatomical kinematic constraint that the human knee frequently acts as a hinge as tested probabilistically.

Chapter 5 contributes a novel method to measure and quantify human crawling motion using four body-worn inertial measurement units attached to the upper arms and thighs through two studies. In the first study, raw IMU data is distilled to four measures of crawling performance; namely, crawl speed, crawl stride time, ipsilateral limb coordination, and contralateral limb coordination. Ipsilateral and contralateral limb coordination quantify the phasing of the rotations of the upper arms and thighs, and thus they are also a product of the underlying motor control (not directly measured). Collectively, these four metrics reveal the purposeful, coordinated movements that contribute to superior (or inferior) crawling performance as further revealed in a cluster analysis that aggregated participants into (statistically significantly different) high and low performance groups. The results of the cluster analysis and statistical tests confirmed the hypothesis that higher performance is associated with limb coordination and stride times. Overall, high performers exhibit superior limb coordination associated with a "diagonal gait" that produces faster crawl speeds and shorter crawl stride times. In contrast, low performers crawl at slower speeds with longer crawl stride times and with significantly less limb coordination. The second of the two studies further demonstrates the ability of these IMU-derived performance metrics to quantify and characterize degradations in human crawling performance as the result of additional body-borne loads. In particular, the results reveal that added load decreases contralateral limb coordination thereby increasing crawl stride time and decreasing crawl speed.

Conclusions

Pivotal to the measurement theory is the anatomical kinematic constraint that the knee can reliably be simplified to a hinge joint when certain criteria are met. Initially, the world frame correction is deterministic and the criteria are fixed to investigate the feasibility of this approach for defining a common world frame. While the theory succeeds during relatively simple movements (i.e., 10 second bouts of low dynamic movements), its performance significantly degrades during more challenging movements (i.e., 2 minutes of exercising the knee analog about all three axes). However, the purely probabilistic measurement theory dramatically improves the accuracy of the IMU estimates for two reasons. First, the criteria for determining when the knee is reliably acting as a hinge joint are developed during functional alignment movements during which this simplification should be valid. For application to human subject testing, this allows for subject-specific probability distributions that characterize that person's (potentially unique) knee kinematics. Second, the probabilistic treatment of the anatomical kinematic constraint injects measurement noise in the Kalman gain calculation. This inclusion of noise in the correction allows for uncertainty in the gross simplification of the knee as a hinge joint. Furthermore, since joint angles are by definition *relative* orientations between body segments, it is therefore sufficient to correct for the *relative integration drift error* between the sensors. This finding is why the measurement theory succeeds without the inclusion of magnetometer data to correct for the *absolute integration drift error* when compared against ground truth provided by the CMM and conventional estimates provided by MOCAP. Future research could explore extending the measurement theory to other joints like the elbow or ankle, both of which may have similar anatomical kinematic constraints that could be exploited for aligning the (independent) sensor world frames.

Next, the results from the systematic survey of IMU anatomical frame definitions show an exponential growth in the number of publications employing inertial motion capture techniques to estimate joint angles. While this subset of publications represents a very small slice of the overall studies employing IMUs, it is reasonable to

assume that this trend is indicative of the growing popularity of the technology in the biomechanics community as a whole, which is significant for two reasons. First, this trend highlights the need to establish a convention for defining anatomical frames of reference for inertial motion capture so researchers can reasonably compare results across studies. Each of the methods revealed in the systematic survey have strengths and weaknesses, but as of today, the most promising method (the self-aligning methods) will require significant validation (and training data) before wide adoption in the research community. Second, the growing popularity also illustrates the importance of developing new (nontraditional) metrics that capitalize on the descriptive power of raw IMU data, much like the work presented in this dissertation for human crawling performance. While there are certainly benefits to conforming to long-standing, position-based performance metrics, there is so much untapped potential with this technology for human biomechanics applications that have never been possible with MOCAP. There are limitless opportunities to apply similar approaches as those used to study human crawling performance to other well-defined tasks by leveraging the strength of IMU technology, namely its ability to directly measure motion.

Part II

Engaged Learning of Engineering Dynamics

CHAPTER 7

Active Learning in Undergraduate Engineering

7.1 Introduction

7.1.1 The Case for Active Learning

In the United States, research on issues concerning teaching and learning gained substantial visibility in 1986 with the Neal Report commissioned by the National Science Board on Undergraduate Science and Education [249]. The first of many reports about the shortcomings of the nation's undergraduate science and engineering education, the Neal Report stressed reforms needed for the U.S. to maintain its "*scientific and technical capacity, its industrial and economic competitiveness, and the strength of its national defense*" [249, p. 1]. As reiterated by the Boyer Commission Report a little more than a decade later, "*Many students graduate having accumulated whatever number of courses is required, but still lacking a coherent body of knowledge or any inkling as to how one sort of information might relate to others...The university has given them too little that will be of real value beyond a credential that will help them get their first jobs*" [250, p. 15]. In response to these and many similar reports, the Engineering Education Research (EER) community answered the call with renewed resolve to improve the quality of undergraduate instruction and launched initiatives to investigate the many facets of teaching and learning. For instance, the National Academy of Engineering (NAE) Engineer of 2020 report [251] and the American Society of Engineering Education (ASEE) Transforming Under-

graduate Education in Engineering initiative [252–255] describe the current state of teaching and learning in engineering and offer insights into necessary improvements and the National Science Foundation (NSF) Revolutionizing Engineering Departments (RED) program [256] facilitates drastic, innovative changes in engineering departments across the nation that enable those improvements.

Stains et al. [257] note that there is a wealth of research reporting other methods like active learning are more beneficial for student learning (see, for example, Prince [30], Hake [258], Michael [259], and Freeman et al. [260]), yet the fact remains that instructional methods emphasizing traditional lecturing still persist as the most common practice in the undergraduate engineering classroom. Borrego et al. [261] found approximately 80% of engineering faculty are aware of at least one active learning innovation, but adoption rates for those innovations are considerably lower. Henderson and Dancy [262] similarly note that achieving sustainable adoption of research-based instructional practices has proven to be extremely difficult.

Recent research has identified several barriers to instructors adoption of active learning that contribute to this slow progress (i.e., the research to practice gap), including insufficient class time, lack of preparation time, and class size (see, for example, Huba and Freed [263], Handelsman et al. [264], Friedrich et al. [265], Froyd et al. [266], Prince et al. [267], and Finelli et al. [268]). Borrego et al. [261], Finelli et al. [268], and Dancy and Henderson [269] all note a common barrier to adoption of active learning is time, both in terms of covering material in class and preparation time outside of the classroom. Henderson and Dancy [262] and Boylan-Ashraf et al. [270] found instructors and researchers also cite the instructors' ability to implement active learning in a large class as a significant concern. This is the core of Benjamin Bloom's "2 Sigma Problem," which describes the need to develop and implement instructional methods for large group instruction that will allow students to perform at the same level they would under individual instruction by a tutor [271].

Motivated by these challenges, the second half of this dissertation introduces and systematically studies an active learning innovation designed for large undergraduate courses in engineering dynamics.

7.1.2 Definitions of Active Learning

A significant body of research supports the benefits of active learning and its overall importance in the future of higher education. In their seminal work, Chickering and Gamson asserted, "*Learning is not a spectator sport. Students do not learn much just by sitting in classes listening to teachers, memorizing pre-packaged assignments, and spitting out answers*" [272, p. 4]. Several years later in a thorough report on the state of active learning in higher education, Bonwell and Eison expanded on that sentiment stating, "*Most important, to be actively involved, students must engage in such higher-order thinking tasks as analysis, synthesis, and evaluation*" [273, p. 5]. This idea of higher-order thinking tasks is founded in Bloom's taxonomy [274–276], which posits that more complex learning opportunities allow students to have more meaningful and impactful interactions with course material.

In Prince's thorough literature review [30], he characterized active learning as anything that allows students to participate in stimulating activities like cooperative learning, problem-based learning, and hands-on exploration to build their conceptual understanding through their own experiences. He investigated how active learning improves conceptual understanding across a variety of engineering disciplines; however, this finding came with the stipulation that these instructional techniques be applied correctly. Prince also drew attention to themes spanning several of the studies included in that review that oppose longstanding assumptions held about engineering education, particularly those related to the tradition of teaching deductively (e.g., starting with a derivation and then providing examples). In a subsequent piece, Prince and Felder [29] reviewed the literature on inductive learning to provide definitions, applications, and supporting evidence for the power of each inductive instructional technique as compared to traditional instructional methods.

Some common types of active learning include inquiry learning, problem-based learning, and experiential learning. Bateman [277] and Lee [278] defined inquiry learning as a technique in which students learn to answer questions, solve problems, or explain observations in such a way that they can think independently and creatively to develop more accountability for their learning and intellectual growth. According

to Prince and Felder [29], project-based learning occurs when a student or a group of students complete an assignment containing one or more tasks that lead to the production of a final product, which could be a design, model, device, or computer simulation. Mills and Treagust [279] reviewed published evaluations of project-based learning in engineering and concluded that students who participate in project-based learning are more motivated, demonstrate better communication/teamwork skills, have a better understanding of issues of professional practice, and know how to apply their learning to realistic problems. Experiential learning, which is directly related to Dewey's experiential learning theory and constructivism, is described by the Association of Experiential Education as "*a process through which the learner constructs knowledge, skill, and value from direct experience*" [280, p. 7]. Gadola and Chindamo [281] noted that experiential learning is generally considered as anything opposite to traditional learning (passive student learners) and is often expressed as 'active learning', 'learning by doing', 'hands-on learning', or any other term that implies an active student learner. A description of this technique applied outside of the classroom is offered by Hajshirmohammadi [282]; specifically, students explored the functionality of different parts of a simulated digital clock, after which they demonstrated their understanding by incorporating a new part into the simulation.

7.1.3 Innovative Pedagogies in Engineering Dynamics Classrooms

Dynamics courses do not typically include opportunities for exploration of the Newtonian mechanics derived and studied in class as noted by Gray et al. [41], Flori et al. [283, 284], Fang and Guo [285], and Fang [286]. As Flori, Koen, and Oglesby succinctly described, "*Dynamics is the study of motion, but textbooks and chalkboards, the traditional classroom teaching tools, cannot show that motion*" [283, p. 61]. Perhaps as a result, engineering dynamics is regarded by some (e.g., Self and Redfield [287] and Rubin and Altus [288]) as one of the more difficult courses in an undergraduate engineering curriculum, likely because of the abstract nature of the subject as noted by Gray et al. [42] and Streveler et al. [289]. Some researchers (e.g., Kozhevnikov et al. [290] and Trindade et al. [291]) hypothesized that students

who explore the physical phenomena related to the mathematical concepts derived in class could be more likely to understand those concepts. To that end, active learning innovations in undergraduate dynamics classrooms can be broadly categorized into one of two pedagogical approaches, namely: 1) computer simulations and/or animations and 2) hands-on experimentation.

The first pedagogical approach integrates computer simulations and/or animations into the curriculum. In addition to improved student conceptual understanding (e.g., Flori et al. [283] or Fang and Guo [285]), a benefit of this intervention is it typically requires minimal additional resources and (once developed) can be easily implemented inside or outside of the classroom. As computers became prevalent in the 1980's, researchers started developing software to visualize the dynamics principles covered in introductory physics and dynamics courses (see, for example, Kraige et al. [292], Gramoll [293], Iannelli [294], and Smith et al. [295]). The most comprehensive example of these early advancements was a 1991 NSF-funded project titled "Dynamic Dynamics" in which Whitman et al. [296] developed software-based modules for particle dynamics concepts that students could engage with and also manipulate. Flori et al. [283] expanded upon this work and showed improvements in student understanding of a small subset of the course concepts covered by their software-based modules for particle dynamics. Stanley [297] explored the student affective response to his software (i.e., how they felt about engaging with the software) and found students believed their overall understanding of the material had improved. More recently, Fang and Guo [285] found that, as a result of engaging with their custom software-based modules, students had significant improvements in development of their conceptual understanding and procedural skills (ability to set up mathematical equations describing the system), and they were able to identify and occasionally correct their misconceptions. Despite their promise, the time commitment required to develop and maintain these software packages makes them difficult to sustain.

Another successful way to engage students in their learning of engineering dynamics is to employ hands-on laboratories during which students conduct experiments and build mental models for specific concepts by observing physical phenomenon

(e.g., Self et al. [298], Adam et al. [299], and Ferri and Ferri [300]). Feisel and Rosa [301] confirmed that demonstrations and experiments can dramatically improve student learning as compared to lecture-only instruction. From a 10-year study that included over 1,000 students, Sharma et al. [302] provided compelling evidence that interactive lecture demonstrations significantly improve student learning gains. Likewise, using data from more than 6,000 students, Hake [258, 303] found students taught with traditional methods had significantly lower learning gains than those taught with interactive engagement methods, which include demonstrations and experiments. In an introductory environmental engineering course, Flora and Cooper [304] found the 109 students participating in a 5-year study favored open-ended experiments to more traditional (plug-and-play) experiments. In particular, the students reported their heightened perceived understanding of course concepts, their increased ability to visualize applications of the theory covered in class, and their improved capacity to create and conduct experiments of their own. However, the large enrollment in introductory dynamics courses often preclude the creation of dedicated hands-on laboratories due to equipment and facility costs [301]. For example, the introductory engineering dynamics course considered in this dissertation (ME240) enrolls approximately 500 students per year. Portable IMU technology (explored in depth in this work) can be used to introduce hands-on laboratories in large classes, thereby providing a feasible alternative to traditional brick-and-mortar laboratories.

7.1.4 Active Learning IMU Intervention

The research in this half of the dissertation includes elements from both the computer simulations/animations and hands-on experiments reviewed above, but with significantly more emphasis on the latter. The active learning exercise allows students to *explore* Newtonian mechanics; thus, it represents an intervention to a teaching pedagogy that otherwise has limited opportunities for student engagement in the learning process. Specifically, the active learning innovation employs inertial measurement unit (IMU) technology (called iNewton) to provide new, hands-on

experiments in an otherwise traditionally taught (i.e., lecture only) introductory dynamics course.

Given the ease with which IMUs yield motion (kinematic) data, they provide a novel and ready-made platform to explore and learn Newtonian mechanics. With carefully designed experiments, students also have the opportunity to simulate motion from derived equations of motion and to compare with what was measured during an experiment. Furthermore, IMUs can be used without expensive laboratory equipment or a traditional (brick-and-mortar) laboratory space that are otherwise hurdles towards adoption. The experiments can be conducted outside of the classroom, providing students with an additional opportunity to meaningfully engage with the course concepts without taking away from in-class lecture time.

7.2 Conceptual Framework and Objective

In the same way that an input to a mechanical system is expected to produce an output, a learning stimulus introduced to a student is expected to elicit a response. In this research, the stimulus is an active learning IMU intervention and there are three responses of interest, namely:

- Conceptual understanding. Streveler et al. [289] defined conceptual understanding in the engineering sciences as including both knowledge about quantities (like acceleration and inertia) and knowledge about the relationships among these quantities (like Newton's 2$^{\text{nd}}$ Law).

- Self-efficacy. Described in more detail in Section 7.3.3, self-efficacy "*refers to beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments*" [37, p. 3].

- Intention to persist. Bean [35] described intention to persist as a students' choice to leave or continue with a field of study. In that study, it was found to have the largest influence on student persistence.

Persistence, retention, and attrition are different, but related behavioral responses frequently used (sometimes interchangeably) in education research and institution assessment. Retention and attrition are measures that evaluate an institution's ability to keep students (retention) or not (attrition) whereas persistence refers to an individual's choice to stay in a specific program, field, college, or institution. Bean's model [35] specifically operationalized student attrition as intent to leave and found that it, along with grades, were the strongest predictors. In this work, student persistence is operationalized as the intent to persist, thereby using it as a proxy for student persistence.

The objective of this study is to test the hypothesis that introducing an active learning IMU intervention will elicit different responses (e.g., conceptual understanding, self-efficacy, and intention to persist) for students who engage with it than for students who do not. As advocated by Svinicki [36], the conceptual framework defined in Fig. 7.1 illustrates the relationship between the active learning intervention and these responses.

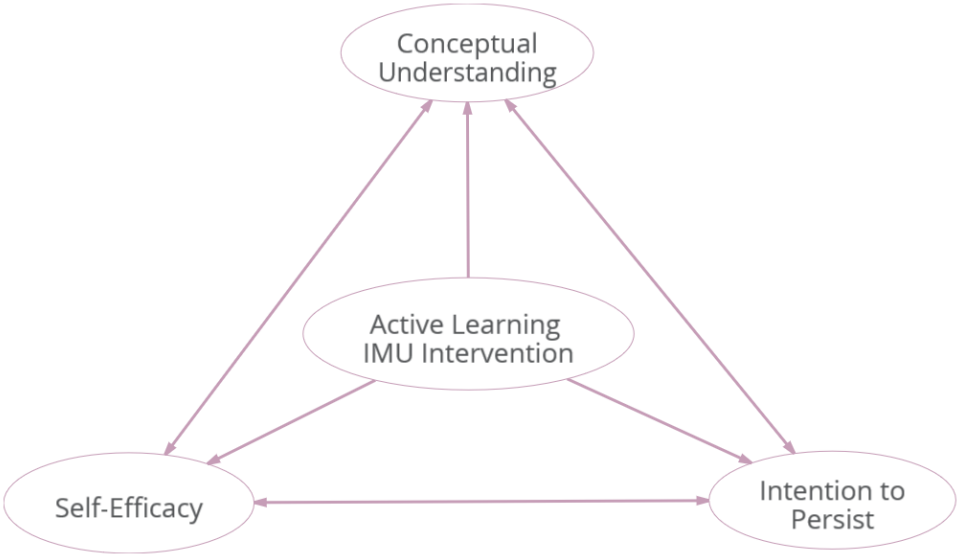


Figure 7.1: Conceptual framework relating the active learning IMU intervention to the three responses (conceptual understanding, self-efficacy, and intention to persist) and relating the three responses to each other.

While there is evidence to suggest that these responses are related to one another, the exact nature of those relationships is currently unknown. Freeman et al. [260] illustrated the relationship between active learning and student performance and failure rates, two related ways to measure *conceptual understanding*. Specifically, they showed improvements in student performance on examinations and decreases in failure rates in core courses in undergraduate programs when active learning was used (students taught with traditional methods were 1.5 times more likely to fail), and they showed active learning has significant positive effects on student performance in small (<50), medium (51-110), and large (>110) class sizes, though the effect size was greatest in small classes. Watkins and Mazur [305] showed that improved academic performance (indicating improved *conceptual understanding*) and increased engagement both play a significant and positive role in retention of students (and therefore students' *intention to persist*) after an introductory physics course. In a similar vein, Graham et al. [306] showed an increase in persistence (and by association *intention to persist*) in college students in STEM disciplines as a direct result of active learning techniques used in introductory courses. Ohland et al. [34] also showed that engagement is a precursor to persistence (and accordingly *intention to persist*) and that institution type had a significant impact in that smaller and/or teaching-focused institutions provided more opportunities for higher levels of learning engagement.

Zeldin and Pajares [307] and Betz and Schifano [33] showed that certain types of active learning exercises, particularly the ones that include hands-on experiences, can improve *self-efficacy*, which has significant implications for selecting career paths, especially for women. Marra et al. [308], for example, showed a strong sense of *self-efficacy* can increase *intention to persist* and ultimately empower students to become practicing engineers, and they showed this to be especially true for women. Brown and Matusovich [309] also found that women generally report lower levels of *self-efficacy* despite comparable academic performance, indicating men and women achieve similar levels of *conceptual understanding*. Vuong et al. [310] showed that *self-efficacy* has a positive effect on GPA and persistence rates (and by association levels of *intention to persist*) for first-generation college sophomores, and Pajares and Miller [311] showed an increase in *self-efficacy* can decrease the level of stress and anxiety

students experience when engaging in an activity. Ballen et al. [312] showed active learning can close the learning gain gap (using *conceptual understanding*-related measures of course grades and knowledge assessment scores) between underrepresented minority (URM) and non-URM students as revealed in a study of more than 250 students in an introductory STEM course. Additional analyses revealed that active learning also increases student *self-efficacy*, which led to the improvements in academic performance for URM students. In an introductory dynamics course that included a variety of active learning methods, Fang [313] demonstrated for 71 students that *self-efficacy* is positively and moderately correlated ($r = 0.55$) with academic performance as measured by average examination scores (a measure related to *conceptual understanding*). Additionally, a thorough review of *self-efficacy* and its implications for academic achievement is documented by Pajares and Schunk [314].

While these three responses (*conceptual understanding*, *self-efficacy*, and *intention to persist*) are clearly related, the focus of the present study is to investigate how different levels of the active learning IMU intervention affects each of these responses individually, rather than to study their relationships. These responses are further categorized as cognitive (*conceptual understanding*) and non-cognitive (*self-efficacy* and *intention to persist*).

7.2.1 Cognitive Responses to Active Learning

Research across several disciplines has provided strong evidence of the cognitive impacts (e.g., students' conceptual understanding) of active learning in STEM courses. Springer et al. [315] conducted a meta-analysis of 39 studies of STEM courses to reveal that active learning in small groups promotes academic achievement. Johnson et al. [316] found in their meta-analysis of 305 studies that cooperative learning promotes higher individual achievement than more competitive or individualistic approaches. Crouch and Mazur [317] reviewed 10 years of studies showing peer instruction in introductory physics courses taught for non-majors improves conceptual understanding and problem-solving abilities. In a controlled experiment involving two introductory undergraduate physics classrooms, Deslauriers et al. [318] found ac-

tive learning increased student attendance, facilitated higher engagement, and more than doubled the learning gains compared to traditional lectures. In large biology courses, Knight and Wood [319] reported significant improvements in learning gains and conceptual understanding in classrooms that incorporated interactive engagement and cooperative work. Freeman et al. [260] also speculate that additional gains in student achievement could be attained through exercises completed outside of the classroom, much like the study design described in Chapter 8.

Research conducted specifically in undergraduate engineering classrooms has similarly demonstrated that active learning has positive cognitive responses. In an engineering statistics course, Kvam [320] showed that active learning methods have a significant impact on knowledge retention for students who scored average or below average on a test. In a nanotechnology engineering course, Hersam et al. [321] iterated on their traditional instructional pedagogy to incorporate collaborative group learning, interdisciplinary learning, problem-based learning, and peer assessment. By comparing final course project scores and course evaluations, they found the increased interactive nature of the class improved student performance by a significant amount while also improving the overall student experience. Bjarne [322] studied an introductory dynamics course in which students anonymously answered clicker questions and subsequently discussed the results with their peers, thereby participating in peer instruction. As compared to a traditionally taught class, the results indicated that this instructional method led to students performing at the same level when solving traditional problems (despite having solved fewer traditional problems in class) while also exhibiting greater conceptual understanding of the subject. In an introductory electrical engineering course, Yadav et al. [94] used traditional lecturing and problem-based learning in an A-B-A-B study design to compare learning gains between the two instructional methods for 55 students. The results suggested the students saw significantly greater (roughly twice) learning gains from the problem-based learning instructional methods than from traditional lecture, though they interestingly perceived greater learning gains from traditional lecture.

7.2.2 Non-Cognitive Responses to Active Learning

Active learning techniques have been shown to be related to non-cognitive responses like self-efficacy and intention to persist. In particular, active learning has been shown to increase students' self-efficacy, directly [323, 324] and indirectly [325, 326]. Fencl and Scheel [323] provide evidence of several types of active learning significantly contributing to student self-efficacy whereas Schunk and Mullen [324] describe how students who are engaged in their learning develop a strong sense of self-efficacy for learning. Similarly, Lent et al. [325] and Hutchison et al. [326] found students report personal (successful) performance experiences as the most common and influential source for their self-efficacy. As is further highlighted by Schunk, "*Sustained motivation depends on students believing that if they change their behavior they will experience better outcomes, valuing those outcomes, and feeling they can change those habits (high self-efficacy)*" [327, p. 72]. As Weiner [328] opines, students persist and achieve at a higher level if they believe added effort will lead to success. For example, Dunlap [329] introduced realistic problem-based learning activities into a senior software engineering capstone course that were intended to reflect the true nature and requirements of the workplace. Using a mixed methods approach, the results indicated that the students' self-efficacy improved (as measured by the student responses on the General Perceived Self-Efficacy Scale), and students further articulated in guided journal entries that they felt prepared to work effectively in their profession.

Hoit and Ohland [330] showed many of the strategies employed to improve persistence in general can significantly reduce the gap in persistence for students historically underrepresented, such as women and students of color. Specifically, they oversaw the conversion of the University of Florida's lecture-based Introduction to Engineering course to one with laboratory activities, resulting in a 17% improvement in retention for the general population with a 36% improvement for women and a 12% improvement for URM students. Lorenzo et al. [331] found that interactive engagement in lecture through cooperative learning significantly reduced the gender performance gap, though they also note that both genders benefited from this ped-

agogical shift. Ballen et al. [312] found a similar result in their study and attributed the improvement in academic performance for URM students to increases in students' science self-efficacy. However, in a follow up study to replicate the results in [331], Pollock, Finkelstein, and Kost concluded, "*Interactive engagement techniques are necessary, but not sufficient for addressing the gender gap*" [332, p. 2]. They indicated that the ways in which instructors enact active learning is critical in how it affects student achievement. The research of Haak et al. [333] also supports the notion that active learning is at least part of the solution to close the achievement gap for underrepresented or underprepared students by providing them with more opportunities to practice their higher-order cognitive skills. In that study, over 100,000 students enrolled in a college-level introductory biology class were included to compare the performance of those who were taught with traditional teaching methods to those who participated in weekly active learning innovations like problem-solving, data analysis, and other activities requiring higher-order cognitive skills.

Active learning has also been shown to improve persistence of STEM students in introductory courses (see, for example, Graham et al. [306]). As Seymour and Hewitt [334] showed in their landmark study, many students opt out of STEM majors as a result of the perceived uninspiring nature of introductory courses that focus more on fundamentals than applications. In response, Felder et al. [335] compared a traditionally taught cohort of 189 students to an experimental cohort of 123 students who took 5 subsequent chemical engineering courses taught with extensive use of active learning. The experimental cohort outperformed the traditional group in measures of retention/graduation in chemical engineering as well as in the percentage choosing to pursue advanced study in the field. However, the authors noted that the success of replicating this approach elsewhere likely depends heavily on the instructor, both in terms of support and experience with implementing these methods. Furthermore, intention to persist has been shown to depend on institution type, perhaps because as Ohland et al. [34] note, smaller institutions usually achieve higher levels of student engagement. Strong evidence arose from their massive aggregated data set (140,000 students), which showed more positive results in faculty interaction, student satisfaction with institution, and overall satisfaction at smaller institutions.

7.3 Learning Theories

To further help ground the conceptual framework in the literature, it is important to anchor it in the context of existing learning theories and three such learning theories are highlighted herein as being relevant to this work. Constructivism is a learning theory which provides the basis for understanding why active learning positively affects student conceptual understanding. The ICAP (Interactive-Constructive-Active-Passive) framework offers a useful way of describing levels of active learning. Lastly, the psychological theory of self-efficacy has significant implications for motivation (e.g., a student's intention to persist in a field of study). Each of these learning theories is reviewed briefly below.

7.3.1 Constructivism

Active learning is fundamentally the realization of constructivism, which has a long history (e.g., Liu and Matthews [336] and Phillips [337]). At the turn of the last century, educational reformer John Dewey, in opposition to behaviorism (the leading learning theory of the time that emphasized rote learning), developed experiential learning theory, which stressed the roles that previous experiences and prior knowledge play in the development of new understanding [338]. Von Glasersfeld [339] described constructivism as a broad learning theory that at its core places the responsibility of learning with the student. Roberts [340] similarly noted, according to that theory, the instructor is viewed as a facilitator whose responsibility is to organize and deliver the content in the form of authentic (real-world) opportunities for structured discovery learning. The success of those learning opportunities "*depends on the quality of the experience which is had*" [338, p. 27].

As is described by Liu and Matthews [336], this student-centered philosophy provided a foundation for two camps of constructivism (each of which has variants of its own): 1) cognitive constructivism and 2) social constructivism. Bodner [341] attributed the origin of cognitive constructivism (sometimes referred to as personal or radical constructivism) to Jean Piaget, who posited that knowledge construction

was not only dependent on previous experiences and knowledge, but that the process occurs in the mind of the learner [342]. Further, as Harlow, Commings, and Aberasturi summarize, "*New knowledge could be constructed only when the learner is confronted with objects (i.e., external experiences) that could not be assimilated into prior knowledge. Rather, the new experience must be accommodated, resulting in a reconstitution of prior knowledge*" [343, p. 45]. In other words, knowledge construction – and by association conceptual understanding – is unique to an individual, which implies that how that knowledge merges with their mental models is also unique.

On the other hand, social constructivism (or realist constructivism), which was guided by Lev Vygotsky's social development theory [344, 345], built from cognitive constructivism to include social interactions with peers and instructors. Individual student learning occurs in a specific community in which those social experiences shape the ways that students think and interpret their world. A community generally has a traditional discourse associated with their conceptual knowledge, which means a physical phenomenon can be understood differently for members of various groups. At the core of both broad categorizations of constructivism is the belief that the students learn most effectively when given the opportunity to personally engage meaningfully and authentically with the concepts.

As noted by Ertmer and Newby [346], learning theories like cognitive constructivism and social constructivism offer guidance for facilitating learning and a foundation for selecting effective instructional techniques like those that fall under the purview of active learning. For example, Rhodes and Rozell described constructivism as providing "*critical guidance on the design and development of activities and materials that create the greatest opportunities for students to learn difficult information, retain that information, and apply it at a later date and even to a novel problem*" [347, p. 174]. They further provided an overview of constructivism as a means for developing student centered electronic texts for an undergraduate physiology course.

Several researchers have drawn from constructivism to design and study teaching approaches in the engineering dynamics classroom. For instance, Bosman et al. [348] followed a constructivist teaching approach by incorporating problem-based learn-

ing activities, on-line discussions, and research/presentations into an introductory engineering course to improve incoming students' communication skills and understanding of basic engineering concepts. Similarly, Gill and Wright [349] successfully developed computer-aided learning activities for students using a constructivist approach with the goal of improved student conceptual understanding of Newtonian mechanics. In a dynamic systems and controls class, Coller [350] designed and implemented a video game-based learning activity with constructivism to make the activities flexible enough that the students have more ownership over the activity without too great a cognitive load. In this current research, constructivism guided the development of an IMU-based learning intervention to reveal and explain the engineering dynamics phenomenon of interest and to interpret the resultant responses (e.g., conceptual understanding and self-efficacy).

7.3.2 ICAP Framework

In applying the theory of constructivism, it is helpful to categorize active learning according to the level of student cognitive engagement. To that end, Chi developed the ICAP framework to describe students' cognitive engagement according to one of four modes (by decreasing level of engagement): *Interactive*, *Constructive*, *Active*, and *Passive* [38, 39]. The underlying hypothesis of this framework is that higher levels of student cognitive engagement result in greater learning gains based on different modes of learning. These modes of learning are largely dependent on four so-called "knowledge change processes," or the dynamic cognitive processes in which students engage while learning new information [39]:

- Store: New information is stored in an isolated way (*Passive*)
- Integrate: New information activates relevant prior knowledge and while storing, new information is integrated with activated prior knowledge (*Active*)
- Infer: New information is integrated with activated prior knowledge, and new knowledge is inferred from activated and integrated knowledge (*Constructive*)

- Co-Infer: Each learner infers new knowledge from activated and integrated knowledge and iteratively infers knowledge with new inputs from conversational partner(s) (*Interactive*)

As Chi and Wiley highlighted in the ICAP framework, the same learning activity can be classified differently based on how the student engages with it [39, p. 228]. For example, Litzinger, Lattuca, Hadgraft, and Newstetter noted, "*Traditional labs where students follow prescribed procedures often fail to engage students in the deep learning processes that lead to conceptual understanding*" [351, p. 136]. Even though labs are designed to guide students through their exploration of a physical phenomenon, students often do not engage with the process as intended.

Many researchers have used the ICAP framework to categorize active learning as a way to compare learning gains. In an introductory biology class, Wiggins et al. [352] used the ICAP framework to develop *active* and *interactive* activities, successfully predicting that improvements in student learning outcomes for *interactive* instructional methods were greater than the gains for *active* instructional methods. Wang et al. [353] used the ICAP framework to develop a coding scheme for categorizing students discussion behaviors in a massive online open course (MOOC) context to explore the relationship between discussion and learning gains. Their results indicated that students who demonstrated *active* and *constructive* behaviors had significantly more learning gains than students who did not, but there were not additional gains for those who engaged in *interactive* behaviors. The authors explain this last finding by noting the necessity that students must be "constructively" *interactive* to reap the benefits of collaboration, with each collaborator contributing meaningfully to the interaction. In further support of this finding and the ICAP framework, Menekse and Chi [354] showed that 72 engineering students' understanding of material science and engineering concepts were significantly greater in the *interactive* condition as compared to the students in the *constructive* condition. The ICAP framework provides a useful way for describing the different levels of learning interventions studied in this work.

7.3.3 Self-Efficacy

Albert Bandura originally proposed and described the psychological theory of self-efficacy as follows: "*Self-efficacy is concerned with judgments of how well one can execute courses of action required to deal with prospective situations*" [44, p. 122]. According to this definition, self-efficacy (a students' realization of their abilities to complete a task) is distinctly different from perceived competence (a students' perception of their ability to complete a task), and Rodgers et al. [355] demonstrated a conceptual and significant statistical difference between these two concepts. The theoretical foundation of this psychological phenomenon illuminates how an intervention can elicit a positive or negative response in developing a person's self-efficacy.

Bandura described four sources for self-efficacy, namely: enactive attainment, vicarious experiences, verbal persuasion, and physiological state [44, 45]. *Enactive attainment* relates to the process of using knowledge, skills, and experience to successfully reach a goal or to complete a task (frequently referred to as a mastery experience). Bandura and his colleagues have shown that success during a mastery experience can improve self-efficacy whereas repeated failures may lower it, especially if failures occur before a resilient sense of self-efficacy can be developed. *Vicarious experiences* are those in which an observer sees others with perceived similar capabilities perform successfully, and this can raise students' self-efficacy with respect to comparable activities. Similarly, observing these others of comparable perceived abilities fail despite their best efforts can lower the observers' judgments of their own abilities. *Verbal persuasion* is overt or covert positive affirmation from others that one can (realistically) master the abilities needed to complete a specific task. With respect to verbal persuasion, Hanson [356] and Hazari et al. [357] showed that the presence or absence of social support can impact overall course achievement. Finally, *physiological state* refers to how mood, emotion, physical reaction, and stress level can influence how one feels about their personal abilities. The most notable example of physiological state is stereotype threat, which Steele and Aronson [358] characterized as debilitating performance anxiety for individuals who identify as a member of a particular group for which there is a negative stereotype associated with the task.

Many studies have used self-efficacy to guide research regarding student academic achievement. In a longitudinal study of 42 students, Lent et al. [359] found students who reported high academic self-efficacy achieved higher grades and persisted more in technical and scientific majors over the following year as compared to those reporting low self-efficacy. In a quasi-experimental study, Bresó et al. [360] developed a self-efficacy-based intervention designed to promote healthy psychological states and engagement in "at-risk" students. Two months after the intervention, the students who participated reported higher levels of self-efficacy and achieved higher levels of performance (defined as the ratio between 'exams taken' and 'exams passed') than those who did not participate. Marra et al. [308] conducted a longitudinal study that included 196 undergraduate women studying engineering at 5 different public institutions. They found significant relationships between the female students' intentions to persist in their current degree programs and their self-efficacy, which partially explains the findings presented by Brainard and Carlin [361] who found more women leave engineering regardless of how their actual performance compared to their peers who persist in the major. In this research, the active learning IMU intervention effectively represents a *mastery experience* in which the students apply their procedural skills and conceptual understanding gained during the traditional instructional methods. Thus, as previous work suggests, the active learning IMU intervention should also build students' self-efficacy and intention to persist.

7.4 Summary

This chapter began by providing evidence in support of the positive cognitive and non-cognitive impacts of active learning across many disciplines of educational research. In response to the active learning IMU intervention investigated in this research, conceptual understanding, self-efficacy, and intention to persist were defined and described in relation to one another and in relation to active learning in general. Finally, the theoretical foundation was described to provide support for the hypothesis that the intervention will incite positive responses in each of these measures. The next chapter describes in detail the intervention and the overall study design.

CHAPTER 8

iNewton Study Design

This chapter provides an overview of the iNewton study design including descriptions of the class setting (and enrollments), a prior pilot study, the iNewton technology, and the overall design of the current study including descriptions of the three levels of the intervention using the iNewton technology.

8.1 Class Setting

The class setting for both a pilot study (described in the next section) and this dissertation's study is an introductory dynamics course, ME240: *Introduction to Dynamics and Vibrations*, required for students (typically sophomores) enrolled in three different programs within the College of Engineering: mechanical engineering, aerospace engineering, and naval architecture and marine engineering. The course also serves as a technical elective for a number of other engineering disciplines. In a typical semester, two or three sections of ME240 are offered, each with an enrollment ranging from 60-120 students for a total annual enrollment of approximately 500 students. The major topics covered in ME240 are three-dimensional (3D) particle motion, planar (2D) rigid body motion, and basic vibrations. The prerequisites include a first-year physics course on Newtonian mechanics and a second-year calculus course on differential equations, which can be taken concurrently. Course grades for the 15-week semester are based on student performance on 13 homework assignments

(15%), two midterm examinations (50%), and a final examination (35%). There is no lab or discussion section associated with this course. Within a semester, each section is taught by a different instructor; however, there are common instructors between semesters. A timeline for a semester is illustrated in Fig. 8.1, which also shows when the active learning IMU interventions are implemented in the class.

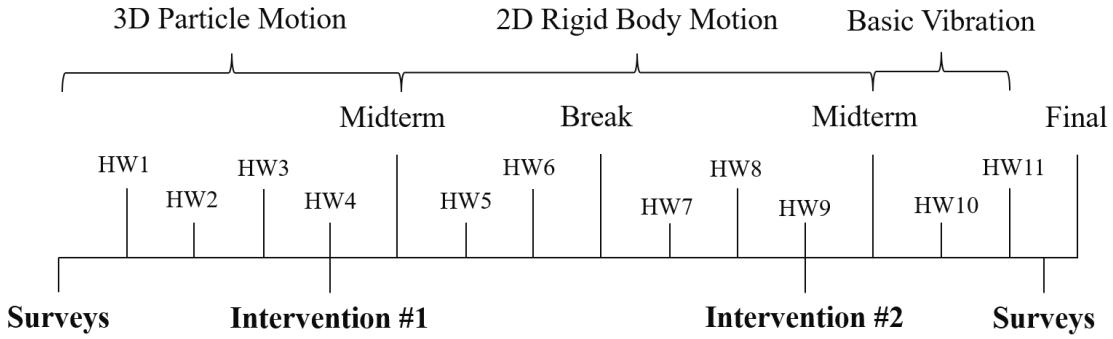


Figure 8.1: Generic timeline for a typical semester. Each tick represents one week, with the survey label at the far left end representing the first week of the course.

8.2 Pilot Study

During the 2013-2014 academic year, a pilot study for the active learning IMU experiments was conducted in ME240 to ascertain the feasibility of implementing this instructional technique in such a large class. This pilot study was funded by a grant from the University's Provost Office for the "Transforming Learning for the Third Century" program, which is dedicated to early-stage education research to increase engaged student learning. Three of the five large sections of the course taught during that academic year served as the control group for the pilot project. The active learning IMU interventions took the form of two instructor-led demonstrations. These were offered in the remaining two sections and comprise the intervention group. Table 8.1 reports the student enrollment for each control and intervention group which, summed over the year, represent a near equal split of the total student enrollment. The two instructor-led demonstrations were conducted during two

Table 8.1: Total number of students enrolled in the pilot study.

	Fall 2013	**Winter 2014**	**Total**
Control	61 (1 section)	174 (2 sections)	235
Intervention	149 (1 section)	72 (1 section)	221
Total	210	246	456

separate lectures (same times as indicated in Fig. 8.1) to highlight concepts in the dynamics of particles and rigid bodies. For example, in one demonstration, the instructor jumped on a pogo stick with an IMU attached to a belt with the instructor modeled as a particle. An in-depth assignment was designed around the data collected from the IMU in which students explored the kinematics and energetics of a particle representing human jumping. The other instructor-led demonstration considered rigid body dynamics and similarly produced IMU data that was passed to the students who used that data to reveal multiple concepts in rigid body dynamics.

The pilot study hypothesized that the demonstrations would positively impact student conceptual understanding, self-efficacy, and intention to persist. Like the current study, student conceptual understanding was assessed using a concept inventory known as the Dynamics Concept Inventory (DCI) [41, 42] (described in depth in Section 9.1.2). The impact on students self-efficacy and intention to persist in engineering was measured using a modified version of the Longitudinal Assessment of Engineering Self-Efficacy (LAESE) [46, 362] (described in depth in Section 10.1.4). Students completed the modified LAESE at the beginning of the term (pre-LAESE) and they completed both the DCI and modified LAESE at the end of the term (post-LAESE). The response rates for the pre-LAESE, the DCI, and the post-LAESE were 81%, 72%, and 54%, respectively of the total number (N=456) of enrolled students.

The overall scores on the DCI showed no significant differences between the control and intervention groups. Students in the Fall 2013 term intervention group did demonstrate greater understanding of the concept related to zero/non-zero acceleration and zero/non-zero velocity, a concept that was addressed in the first of the two instructor-led demonstrations that term. For this concept on the DCI, the scores were

statistically significantly higher in the intervention group (53±3.9% vs. 26±6.0%, p<0.0001). For most LAESE items, there were no significant differences between the control and intervention group, but gains in students' course-specific self-efficacy for the Winter 2014 term was significantly higher in the intervention group (2.5±1.2% vs. 7.6±2.1%, p<0.001). These early findings warranted a thorough investigation, which is captured by the study completed in this dissertation. The design of that current study is described in the remainder of this chapter.

8.3 Participant Demographics for Current Study

Data were collected from 1,070 of the 1,388 students enrolled in 19 sections of ME240 over 3 years, for an overall response rate of about 80%. The timeline and enrollment details for the study are illustrated in Fig. 8.2 below. Demographic data was collected with the beginning of the semester survey and is displayed in Table 8.2. Students were incentivized to complete the surveys with modest course extra credit for completion (not performance). To discriminate between students who completed the survey questions with effort from those who did not, three inclusion criteria were used: 1) number of questions answered, 2) time spent taking the survey, and 3) longest run of the same answer (e.g., selecting the response "a" repeatedly). Furthermore, there were also students who took the class multiple times thereby completing the surveys on more than one occasion. For these cases, the first instance of their participation was retained in the study. Finally, data from students who only completed one of the two surveys were eliminated as well. As a result of these exclusions, the final data set included 1032 students for a response rate of 75%.

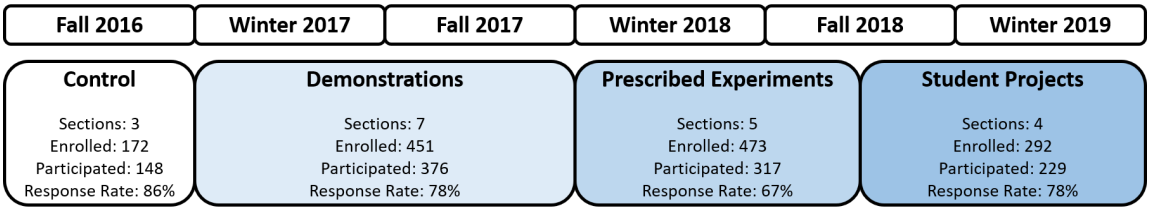


Figure 8.2: Timeline and enrollment data for the study.

Table 8.2: Demographic data for students that completed the surveys at the beginning of each of the indicated semesters. Level 1 is the demonstrations group, Level 2 is the prescribed experiments group, and Level 3 is the student projects group.

	Control	**Level 1**		**Level 2**		**Level 3**
	FA16 N = 148	**WN17** N = 244	**FA17** N = 109	**WN18** N = 197	**FA18** N = 127	**WN19** N = 207
Gender						
Male	128 (86%)	181 (74%)	84 (77%)	141 (71%)	105 (82%)	159 (77%)
Female	21 (13%)	60 (25%)	21 (19%)	55 (28%)	22 (17%)	48 (23%)
No Response	1 (1%)	3 (1%)	4 (1%)	1 (1%)	0 (0%)	0 (0%)
Ethnicity						
White, Not of Hispanic Origin	82 (56%)	166 (68%)	74 (68%)	129 (65%)	83 (65%)	125 (60%)
Black/African American	5 (3%)	7 (3%)	2 (2%)	3 (2%)	2 (2%)	8 (4%)
Hispanic/Latino	2 (1%)	12 (5%)	2 (2%)	13 (7%)	4 (3%)	11 (5%)
Asian	48 (33%)	43 (16%)	18 (17%)	35 (17%)	21 (16%)	41 (20%)
American Indian/Alaskan Native	0 (0%)	0 (0%)	1 (1%)	0 (0%)	1 (1%)	0 (0%)
Native Hawaiian/Pacific Islander	0 (0%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Two or more	8 (5%)	7 (3%)	9 (8%)	10 (5%)	8 (6%)	15 (7%)
Unknown/Do not wish to report	2 (1%)	4 (2%)	3 (2%)	5 (3%)	6 (5%)	6 (3%)
No Response	1 (1%)	4 (2%)	0 (0%)	2 (1%)	2 (2%)	1 (1%)

8.4 Intervention Levels for Current Study

The active learning IMU intervention has three levels (Table 8.3) that systematically increase student engagement with the IMUs (incremental assessments detailed by Vitali et al. [363, 364, 365]). These three levels, referred to as 1) Demonstrations, 2) Prescribed Experiments, and 3) Student Projects, respectively map onto the *passive*, *active*, and *constructive* modes defined by Chi's ICAP framework [38, 39]. Students watch instructor-led demonstrations as passive observers in Level 1, they actively conduct prescribed experiments in Level 2, and they construct their own experiments to study specific concepts from class in Level 3.

Table 8.3: Descriptions of the intervention levels with the IMU-based experiments.

Level	**Intervention**	**Description**
0	Control	Instructors teach course without intervention
1	Demonstrations	Instructors demonstrate experiments with IMUs in class
2	Prescribed Experiments	Students conduct prescribed experiments with IMUs outside of class
3	Student Projects	Students propose and conduct an experiment of their own imagining (with instructor feedback) with IMUs outside of class

Misunderstood concepts identified from the DCI for the control group formed the basis for designing the demonstrations and prescribed experiments. For the student projects, students designed experiments concerning concepts in planar rigid body dynamics. A detailed description of each of the three intervention levels is provided next. Sample assignments for each of the intervention levels are provided at `https://inewton.engin.umich.edu/`. The following two chapters discuss the effects these intervention levels have on student conceptual understanding (Chapter 9), self-efficacy (Chapter 10), and intention to persist (Chapter 10).

8.4.1 iNewton IMUs

The IMUs (Fig. 8.3) used and provided to the students to conduct the experiments, and known as iNewtons, were custom made specifically for this project (Insight Limited, Hong Kong).

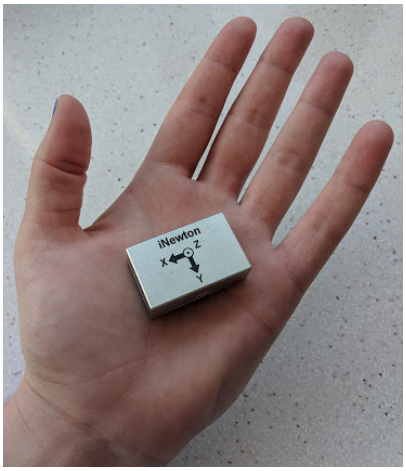


Figure 8.3: Photo of the iNewton IMUs used by the students in this project.

Each sensor contains a triaxial accelerometer and triaxial angular rate gyroscope with programmable ranges and sampling frequencies as follows:

- Accelerometer: ±4g, ±8g, ±16g, ±32g
- Angular Rate Gryoscope: ±500°/sec, ±1000°/sec, ±2000°/sec, ±4000°/sec
- Sampling Frequency: 100 Hz, 200 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz

Turning the sensor on (sliding the switch to the "on" position) immediately starts writing acceleration and angular rate data to on-board flash memory. The data collection ends when the iNewton is turned off (sliding the switch to the "off" position). The data can then be downloaded to a host computer via USB where the iNewton is recognized as an external storage container. Further details on the operation of the iNewton are provided in the User Manual that is accessible at `https://inewton.engin.umich.edu/`, which was also provided to the students.

8.4.2 Level 1 Intervention: Demonstrations

Two demonstrations were conducted in class by instructors, and they each took no more than 15 minutes to set up, conduct, and explain the relevant concepts revealed by the experiment. The first demonstration (Fig. 8.4) focused on measuring and understanding the Coriolis acceleration in the context of particle dynamics. The IMU was attached to a slider (the particle) that was free to slide along a rotating arm. An approximately constant force was applied to a string to generate a constant moment on a shaft that rotated the arm. The rotating shaft was rigidly attached to the arm, thus the constant moment created a constant angular acceleration of the arm. The experiment included the phase of motion where the slider stuck to the arm, followed by the phase when it slid outwards along the arm (ultimately impacting a hard stop). Students modeled this experiment by deriving the equations of motion of the slider (with the attached IMU) as a particle and were asked a series of conceptual questions to be answered using the recorded IMU data.

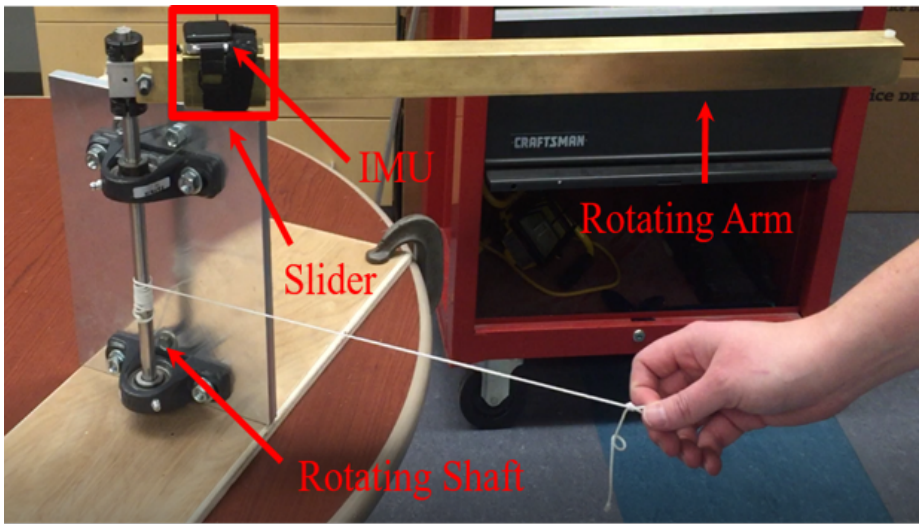
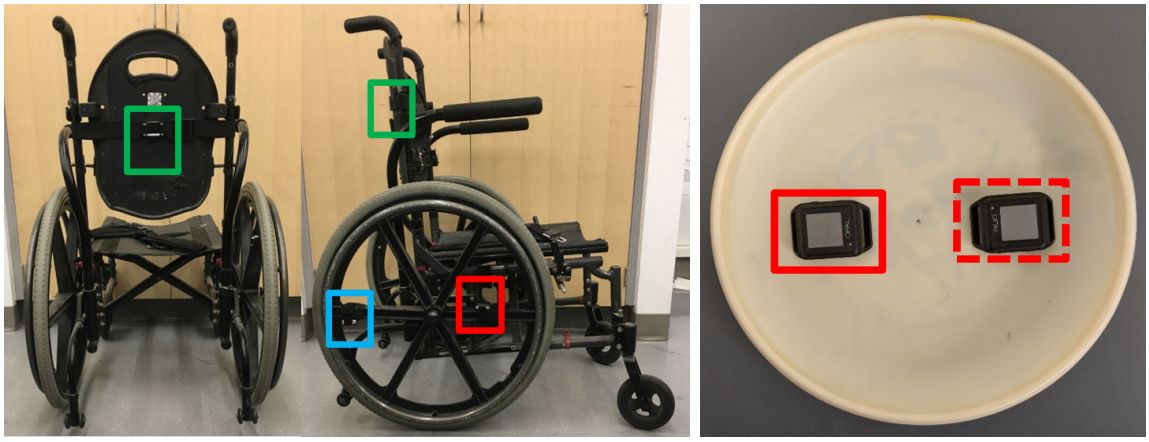


Figure 8.4: Experimental set-up of a rotating arm with a slider that demonstrates the Coriolis acceleration (Demonstration #1).

The second demonstration was designed to study rigid body kinematics, rolling without slipping, and Newtons 2nd law for a rigid body. Two versions of this demon-

stration were offered in different semesters. The first version (Fig. 8.5a), demonstrated in Winter 2017, consisted of three IMUs attached at three locations on a wheelchair; namely, on the outer perimeter of a wheel, near the axle of the same wheel, and on the back of a chair. The second version (Fig. 8.5b) of the demonstration, used in Fall 2017, consisted of two IMUs attached at radially-symmetric locations on the underside of a Frisbee. In both versions, the object was pushed to produce initial linear and angular velocities that then allowed to roll freely thereafter subject to dissipative effects. Following the demonstrations, students were given the relevant (pre-processed) data collected with the IMUs to complete an assignment designed to expose and explain these concepts.



(a) (b)

Figure 8.5: The two versions of Demonstration #2. (a) The wheelchair version included three IMUs located on the back of the chair (green), on a wheel near the outer perimeter (blue), and on the same wheel near the axle (red). (b) The Frisbee version included two IMUs located symmetrically on the underside. The IMU in the solid red box collected data for the assignment whereas the IMU in the dashed red box was placed to minimize the effects of an eccentric mass on a rotating object.

Follow-On Assignments for Demonstrations

For the first demonstration, students estimated the angular acceleration of the rotating arm from a linear approximation for the recorded angular velocity. Then, they drew a free body diagram of the slider when it was sliding along the rotating

arm. Using Newtons 2nd law, they developed an equation of motion for the slider in the radial direction that they solved numerically for the slider velocity and position. These results were used to estimate the normal lateral force acting on the slider, which has components deriving from the Coriolis acceleration and the arm angular acceleration. They then compared the force component magnitudes, which reveals the Coriolis component is much larger than the angular acceleration component.

The first version of Demonstration #2 (Fig. 8.5a) included a portion focused on the rigid body kinematics of the wheel. Students compared the angular velocities recorded by the IMUs at the two positions on the wheel to confirm (or to learn) that they were the same. Then, they resolved the components of the recorded acceleration from one IMU in the sensor frame of the second IMU to confirm (or to learn) that the accelerations were different in both magnitude and direction, but also similar in their (out of phase) fluctuations. A second portion of the assignment focused on rigid body kinetics. This portion required drawing a free body diagram to understand how to estimate the force from the push at the beginning of the trial as well as the dissipative rolling resistance force after the initial push ended. For the second version of Demonstration #2 (Fig. 8.5b), students estimated the velocity of the center of mass of the Frisbee using only the measured angular velocity, knowing that the Frisbee was rolling without slipping. Using this result, they further computed the translational and rotational kinetic energies of the Frisbee. From the work-energy principle, they estimated the work done by the dissipative forces and also specified which forces acting on the Frisbee did work.

8.4.3 Level 2 Intervention: Prescribed Experiments

The prescribed experiments for the Level 2 intervention were performed by the students outside of class, and so they were designed to be conducted anywhere with an easily transported experimental set-up. Furthermore, students worked in groups of two or three to encourage collaborative learning and to minimize the number of experimental set-ups. Prescribed Experiment #1 (Fig. 8.6) focused on familiarizing the students with the iNewtons as well as developing student intuition for accelera-

tion. The goal for this experiment was to balance an inverted pendulum on the palm of a hand and use the measured acceleration from the iNewton to understand the dynamics of active balancing treating the inverted pendulum bob as a particle.

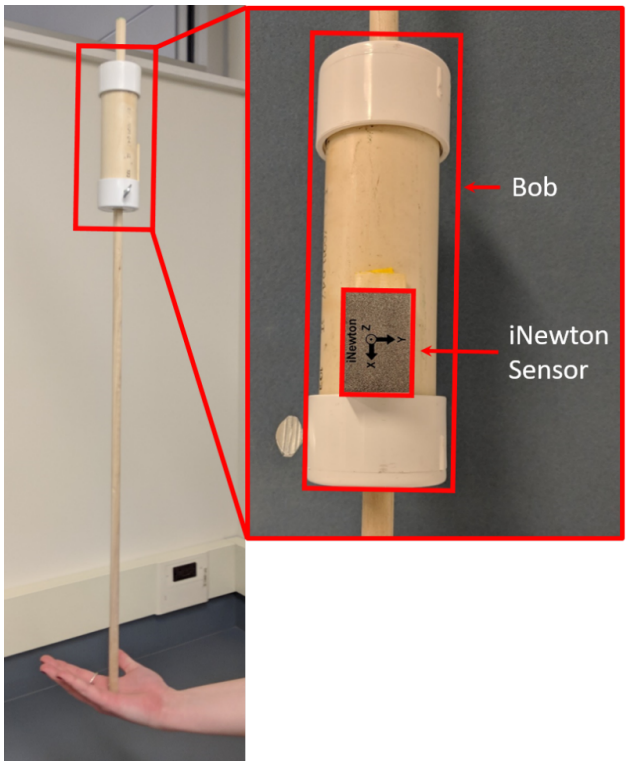


Figure 8.6: Photo of the experimental set up for Prescribed Experiment #1 with a callout of the iNewton attached to the bob of the inverted pendulum.

The assignment was purposefully kept simple so as not to overwhelm the students with conducting the experiments on their own and how to interpret data from the iNewton. Students used the fact that the iNewton remained essentially still and so the acceleration data described gravity and thus also the tilt of the pendulum from vertical. Students also explored the difficulty of balancing the pendulum under a variety of conditions (e.g. using dominant versus non-dominant hand, pendulum bob high above hand versus closer to hand, and both eyes open versus one eye closed).

In the first version offered in Winter 2018, students studied their ability to balance

the inverted pendulum at three different predefined heights. They were asked to repeat the experiment with the bob at the tallest height with both hands. Whichever hand did better (using a balance performance metric they estimated from iNewton acceleration data), they were then instructed to conduct the experiment two more times with the bob at the other two (more difficult) heights. In the second version offered in Fall 2018 and Winter 2019, students studied their ability to balance the inverted pendulum at the tallest height with their dominant hand with both eyes open. They then compared that to their performance balancing the pendulum with one eye closed, effectively eliminating their depth perception and making the task considerably more difficult.

Prescribed Experiment #2 (Fig. 8.7) follows the design of Demonstration #2 with a Frisbee. Consequently, it was designed to study rigid body kinematics, rolling without slipping, and Newtons 2nd law for a rigid body. Students were instructed to

Figure 8.7: An example configuration for the iNewton sensor attached to the Frisbee (Prescribed Experiment #2).

push the Frisbee in such a manner as to produce initial linear and angular velocities and then allow it to roll freely thereafter subject to dissipative effects. They measured how far the Frisbee rolled and compared it to the distance they estimated from the iNewton data.

Following the experiments, the students were given code to pre-process the data collected with the iNewtons and to complete an assignment designed to expose and explain the relevant course concepts.

Follow-On Assignments for Prescribed Experiments

For the first prescribed experiment, students were asked to derive the lean angles of the pendulum relative to the horizontal plane (i.e. pitch and roll) using the components of acceleration measured by the iNewton. They used these angles to estimate the coordinates of the pendulum bob in the horizontal plane relative to their hand using the known height of the bob. For the first version of the experiment, the students computed and plotted the location of the bob after each of the four experiments (three with their "better" hand and one with their "worse" hand). Then, they calculated the root-mean-square (RMS) of the radial position, which is a measure of balance performance. They compared balancing performance across the four trials. For the second version of the experiment, students also computed and plotted the location of the bob for both trials (both eyes open and one eye closed) and similarly calculated and compared the RMS of the radial position. They were also asked to model their control strategy for keeping the pendulum upright as a spring (effectively describing a proportional controller). To this end, they numerically differentiated the Cartesian coordinates of the bob twice to obtain acceleration and, from that, estimated the horizontal force applied to the bob (via Newton's 2^{nd} law). A scatter plot of this force plotted against position reveals an approximately linear relationship, and the slope of the fitted line reveals the gain employed in their control strategy (i.e. the stiffness of an effective spring).

For the second prescribed experiment, students studied both the kinematics and kinetics of the rolling Frisbee. They estimated the velocity of the mass center using the angular velocity data. Numerically integrating this velocity estimate produced

an estimate for the horizontal position of the Frisbee, the final value of which they compared to the distance they measured during the experiment. Numerically differentiating the estimated velocity produced a noisy estimate for the acceleration of the Frisbee, which they compared with a (low pass) filtered estimate from a template code. Using their estimate for mass center velocity and the measured angular velocity, students then calculated the kinetic energy of the Frisbee. They also identified the likely sources of energy dissipation and used the work-energy relationship to determine the work done by each source. A challenge question was also posed, which most students attempted, in which they derived expressions for the friction forces acting on the Frisbee and the work done by each of those forces. This question highlighted the fact that with the rolling without slip assumption the friction force between the Frisbee and the ground does no work because there is no relative motion.

8.4.4 Level 3 Intervention: Student Projects

For this version of the intervention, the students worked in groups of two or three to complete one prescribed experiment early in the term and then one student project near the end of the term. The prescribed experiment was offered first to allow students to learn how to use the iNewton prior to designing their own experiments. The prescribed experiment was the same inverted pendulum experiment described in Level 2 in which they compared their balancing performance with both eyes open to their balancing performance with one eye closed.

The student projects provided a unique opportunity to students to conceive of their own experiment and to pursue a study in rigid body dynamics that interested them. This project was broken into three smaller parts to promote progress and provide students with timely feedback. In the first part, the student groups proposed two or three potential experiments using the rubric provided in Fig. 8.8 to guide them. Students were requested to list of the course concepts in rigid body dynamics revealed by the experiment, describe a proposed experimental procedure, predict what they expect to see in the data measured by the iNewton, and discuss any potential risks in conducting their experiment. Instructors then evaluated the

proposals to make sure at least one of the experiments was achievable and acceptably safe for the students to pursue.

In part 2, the groups submitted a preliminary report that showed progress towards one of the instructor-approved experiments and discussed any challenges they were facing. The preliminary report included a description of the experimental procedure used (including a video showing the procedure) with special attention paid to how the iNewton was oriented on the rigid body of interest. They provided plots of their preliminary measurements along with a discussion of how these measurements compared to their predictions made in the first part of the project. The reports also included a preliminary analysis in which they developed free body diagrams and equations of motion needed to derive specific kinematic or kinetic variables of interest. Finally, students reiterated any risks associated with their chosen experiment. At

	Conceptual Difficulty		
	Minimum	Expected	Advanced
System Chosen for Study	A particle moving in 2D	A rigid body moving in 2D	A system of two (or more) rigid bodies moving in 2D
Model and Analyses	Develop a model of your system by drawing a free body diagram(s). If appropriate, derive the equations of motion (Newton's 2nd Law and/or Euler's 2nd Law)	Explore additional kinematic or kinetic variables that further explain the motion drawing from additional concepts, including velocity and/or acceleration analyses (Ch. 5) and/or work-energy (Ch. 6).	
Visualization[1]	Plot raw iNewton data as functions of time	Plot the additional kinematic or kinetic variables derived from iNewton data	Animate your system's motion by any means, including (if appropriate) by solving the ODEs
Interpretation and Reflection	Describe the course concept(s) from Chapters 5 and/or 6 revealed in your experiment and how the model helps explain the motion you observe	Describe how the additional kinematic or kinetic variables help explain the motion you observe	Describe how the animation helps explain the motion you observe

Figure 8.8: Rubric describing the general expectations for the student projects.

this point, instructors offered feedback including alternatives to the experimental procedure and/or proposed analyses to better reveal course concepts.

In part 3, students incorporated the feedback offered by the instructors into their projects and completed a summary report. The summary report included an introduction listing the course concepts and reasons for the group's decision to pursue their experiment. They then described the final experimental procedure, analytical model, and analyses. Final results and a discussion comparing their experimental data with their predictions were included as well. Lastly, students reflected on the course concepts they elected to study and described whether the experiment they designed and conducted reinforced their understanding of these concepts.

8.5 Chapter Summary

To summarize, this chapter provides a detailed description of the study design including the class setting, a pilot study, participant demographics for the current study, and descriptions for each active learning IMU intervention level. The next two chapters will discuss the effects of the active learning IMU intervention levels on student conceptual understanding (Chapter 9), self-efficacy (Chapter 10), and intention to persist (Chapter 10).

CHAPTER 9

iNewton Cognitive Effects

9.1 Background to Studies

As described in detail in Chapter 7, engineering dynamics is historically challenging for students to understand and to transfer concepts to new contexts as demonstrated by Gray et al. [41], Flori et al. [283, 284], Fang and Guo [285], and Fang [286]. It is especially difficult for first-time learners to imagine motion through static examples and illustrations [283, 285], which are common in traditional instructional methods (e.g. lecturing, note taking, and book problem solving) typically only passively engaging students with the material. Pertinent to this dissertation are active learning innovations that employ inertial sensors (either self-contained IMUs or smart-phone-embedded sensors) for active learning as reviewed next in the context of various engineering courses and, more specifically, undergraduate dynamics courses.

9.1.1 Educational Studies Using IMUs

Given the versatility of IMUs, researchers have utilized this technology for a wide range of educational applications and disciplines. For example, Espinosa et al. [366, 367] employed inertial sensors as a means to teach human biomechanics. Specifically, undergraduate distance learners used IMUs to conduct biomechanical experiments to complement their theoretical curriculum. At the United States Naval

Academy, Bradshaw and Nicholson [368] provided inertial sensors to systems engineering students to facilitate learning principles of navigation systems (including coordinate transformations and Kalman filtering) in a prescribed experiment. Brill et al. [369] designed a laboratory for an undergraduate controls course in which students controlled the position of an actuated arm using outputs provided by a smartphone application. A follow-up expert analysis and a content knowledge assessment revealed this activity yielded significant improvement in understanding of controls concepts. The authors further expand on how smart-phone platforms are feasibly implementable for a wide range of laboratory education for science and engineering fields [370]. Chen et al. [371] describe a senior capstone design project wherein students utilized a wearable, wireless triaxial accelerometer to monitor human motion. Similarly in a design course, Jordan and Lande [372] highlight a student group who visualized arm movements from data collected by an IMU embedded in a bowling ball. In an upper division mechanical engineering laboratory course, Nordenholz [373] provided dual axis accelerometers and uniaxial angular rate gyroscopes to students who investigated the planar motion of remote controlled cars and estimation errors. Finally, Shayesteh et al. [374] incorporated wearable and Internet of Things (IoT) devices in a senior course to introduce various engineering applications via hands-on experiences to mechanical engineering and electrical and computer engineering students, who responded positively to the experience.

Several studies of implementing IMU technology as a learning platform have also taken place in undergraduate introductory dynamics courses as well. Bevill and Bevill [375] designed a sophomore-level engineering dynamics laboratory activity in which students compared normal and tangential components of acceleration measured by an accelerometer embedded in a smartphone to those derived from GPS data measured by the smartphone. Students reported enjoying the process of collecting and analyzing data outside of a laboratory setting and feeling that the activity contributed positively to their conceptual understanding of particle dynamics. The authors also note they designed this experiment specifically because students tend to struggle with kinematic analyses [375], one of the cornerstones of Newtonian mechanics as evidenced by its prevalence in the Dynamics Concept Inventory [41, 42]

(described subsequently) and the Force Concept Inventory [376]. Ferri and Ferri [300] describe several different hands-on experiments, including those facilitated by IMUs, to provide additional learning opportunities in an undergraduate dynamics course. Students reportedly prefer and benefit from conducting experiments, which the authors attributed to the fact that experiential learning activities (i.e., hands-on experiments) can influence memory and recall. Finally, students designed and conducted experiments of their own imagining with their smartphones, which resulted in a variety of different projects ranging from studying Coriolis acceleration to studying the kinematics of a home-made trebuchet [377]. O'Connor reported "*overwhelmingly positive feedback from the student survey, as well as the remarkable nature of the reports, this first student-led investigation project into dynamics theory was an absolute success*" [377, p. 12].

Given the recent history of employing IMUs for teaching undergraduate engineering dynamics, it is important to measure the impact of this new form of engaged learning on student conceptual understanding, which was only formally assessed by Brill et al. [369]. As described by Treagust [378] and Fulmer et al. [379], diagnostic evaluation of student conceptual understanding takes a number of forms in student outcome assessment and education research. Researchers like Osborne and Gilbert [380], Nelson [381], and Watts [382] advocate the use of interview sessions with students to probe their understanding of the subject matter, though this approach is not scalable to large classes (as in this research). Others like Kirbulut and Geban [383] and Tamir [384] opt for multiple choice questions populated with answers usually developed from students past answers and typifying frequently misunderstood topics known as distractors. Concept inventories (CIs) are an example of the latter. As Jorion et al. [43, 385] describe, CIs undergo extensive evaluation to assess validity and reliability in assessing student conceptual understanding and ensuring that similarly-grouped items test the same underlying concept. This will be discussed further in Section 9.2.2. Furthermore, CIs like the Dynamics Concept Inventory [41, 42] have been used in both high-stakes assessment (e.g., counted in the score on a final exam) and low-stakes assessment (e.g., offered outside of the normal course grading scheme). However, as is noted by Finn [386], scores from low-stakes evaluations may

not produce unbiased representations of student knowledge, which will be discussed in Section 9.2.1. Next is an introduction and description of the Dynamics Concept Inventory.

9.1.2 The Dynamics Concept Inventory

At a mini-conference on Undergraduate Education in Dynamics, Vibrations, and Strength of Materials in 2002, attendees discussed instructional innovations in the physics education community that might benefit engineering mechanics as well [41]. The attendees concluded that a new assessment tool would first be needed to evaluate the efficacy of future teaching innovations for Newtonian mechanics [42]. Headed up by Gray et al. [41, 42], the Dynamics Concept Inventory (DCI) was borne from that initiative. The DCI consists of 29 questions focused on 14 important and/or commonly misunderstood concepts in engineering dynamics [385]. The concepts, provided in Appendix B, were identified via a modified Delphi process by veteran course instructors [42].

A number of prior studies employ the DCI to evaluate potential gains in conceptual understanding following changes in teaching pedagogy in undergraduate dynamics courses. Shelley [387] used the DCI as an objective metric for iteratively altering teaching instruction for improving student learning outcomes. Benson et al. [388] used the DCI to quantify improvements from the Student-Centered Activities for Large Enrollment Undergraduate Programs (SCALE-UP) approach, which incorporated inquiry-based learning exercises. Similarly, Self et al. [298] used the DCI to assess the effects of two inquiry-based learning activities that aim to correct student misconceptions concerning friction forces and the dynamics of rolling bodies. Employing the DCI, Bedillion et al. [389] evaluated the effects of including SolidWorks motion simulations to improve visualization skills and knowledge transfer. Tang and Bai [390] used the DCI to evaluate the effectiveness of identifying elements of problem solving skills, developing exercises to improve those elements, and creating sequentially more complex activities for improved transferability.

In 2015, Jorion et al. determined the "*DCI can function as a low-stakes instru-*

ment that educators can use to identify overall understanding of all concepts identified in the DCI" [43, p. 479]. Similarly, Stites et al. [391] described a reliable and valid abbreviated DCI that correlates with student performance on traditional long-answer problem-solving exam questions in a high-stakes setting. These two studies conducted by Jorion et al. [43] and Stites et al. [391] provide support for the reliability of the DCI as an evaluative tool to measure student conceptual understanding and represent the basis of the first study described in Section 9.2.1. However, the findings from the 2015 study [43] also raised concerns about the DCI that are further discussed in Section 9.2.2.

9.2 Three Studies Concerning Cognitive Effects

The remainder of this chapter is organized into three studies that all probe the conceptual understanding of students in ME240. The first study considers the authenticity of student performance on the DCI in a low-stakes setting. The second study considers the validity and reliability of the DCI as an instrument designed to measure student conceptual understanding. The third and final study uses the results from the first two studies to investigate the impact of the active learning IMU intervention on conceptual understanding.

9.2.1 Study 1: Authenticity of DCI Performance in Low-Stakes Settings

Study Background and Objective

This first study investigates how conceptual understanding in a low-stakes setting is related to both conceptual understanding and performance on traditional long-answer problem-solving in a high-stakes setting. A "high-stakes" setting is one in which there are significant consequences for students' grades (e.g., a midterm examination) while a "low-stakes" setting is one in which there is little grade consequence (e.g., an assignment for extra credit independent of their performance). In this study, the DCI is offered in a low-stakes setting, namely as part of an online survey at both the beginning (pre) and end (post) of the semester. Students are

incentivized to participate with modest course extra credit for filling out any portion of the survey (independent of performance). Given the leniency of these testing conditions, the first hypothesis concerns whether student performance on the low-stakes evaluation (DCI post) will correlate with that on two high-stakes evaluations, namely short-answer concept questions and long-answer problems on in-class examinations. It is hypothesized there will be low correlation between performance on the low- and high-stakes evaluations. It should be noted that prior studies conducted by Ates and Catologlu [392], Malone [393], and Stites et al. [391] show conceptual understanding can be correlated with problem-solving skills. Therefore, a secondary hypothesis is there will be a statistically significant correlation between the student scores on the two high-stakes performance assessments (i.e., the short-answer concept questions and the long-answer problems). As a result of the anticipated correlation, a third hypothesis is that student performance on the two high-stakes evaluations will be predictive of their performance on the low-stakes evaluation.

Methods

The DCI was offered twice during one term (Fall 2017) for one course section near the start and end of the term (weeks 2 and 15). Students took the DCI in the low-stakes setting of an online survey (see Fig. 8.1 for a semester timeline). The online survey was offered without time limitation, and students who completed any portion of the DCI received modest course extra credit (e.g., 1% course extra credit). For each administration, the performance of each student is calculated as the percentage of the 29 questions answered correctly. For the subsequent analysis in this first study, only student performance on the DCI administered at the end of the semester is considered. Fifty-seven of the 70 students enrolled in this section of the course participated (81% response rate).

A midterm examination was administered 12 weeks into the 15-week semester that focused on concepts for the unit on rigid body dynamics. That examination included six questions, namely: 1) a set of four short-answer concept questions related to the course concepts, and 2) two traditional long-answer problem-solving questions. The 4 short-answer questions were formulated around 6 of the 14 concepts on the DCI.

As an example of a short-answer concept questions, students were asked to identify the relationship between angular velocities measured at two points on the same rigid body. One of the long-answer problem-solving questions probed the same concept by asking students to complete velocity analyses for a wheel rolling without slipping.

Each long-answer problem-solving question was graded by the same person (either the instructor or graduate student instructor) for consistency. Performance on this examination contributed to 25% of the students final course grade, representing a high-stakes evaluation of concepts in rigid body dynamics. This examination was offered at the conclusion of the unit on rigid body dynamics, and thus may coincide with the peak in student understanding of this material.

Due to the nonnormalities in the data, Spearman correlation coefficients were calculated to evaluate the first two hypotheses regarding the relationship between student performance on the low-stakes evaluation (DCI post) and the high-stakes evaluations (short- and long-answer questions on midterm examination). To address the third hypothesis that high-stakes performance is predictive of low-stakes performance, stepwise linear regression (backwards elimination) was conducted to determine a final model with the best fit. The predicted (outcome) variable is the DCI score at the end of the semester (DCI post), and the regressors are the scores on the high-stakes midterm evaluation (short- and long-answer scores). Normality and heteroscedasticity of the residuals are confirmed for the final model.

Results and Discussion

Table 9.1 documents student performance on the DCI and summarizes the low-stakes evaluation of student conceptual understanding. Reported are the mean (standard deviation) of overall DCI scores taken at the end of the term (post). Table 9.1 also documents student performance on the midterm examination, the high-stakes evaluation of student conceptual understanding. Reported are the mean (standard deviation) of the four short-answer concept questions and the two long-answer problem-solving questions used to evaluate student conceptual understanding.

Table 9.1: Means (standard deviations) for low-stakes (DCI) and high-stakes (Short- and Long-answer questions on midterm examination) evaluations of student performance.

	DCI %	**Exam Short %**	**Exam Long %**
Scores	48.0 (19.7)	75.3 (19.1)	84.5 (13.2)

To evaluate the relationships between the scores, Spearman correlation coefficients (r_S) are calculated for each pair and are reported in Table 9.2.

Table 9.2: High-stakes evaluation of student performance.

		High-Stakes	
		Exam Short	Exam Long
High-Stakes	Exam Long	0.68***	
Low-Stakes	DCI	0.42**	0.43***

Significant at α = *0.05, **0.01, ***0.001.

The first hypothesis is supported by the moderate correlation between the DCI post scores and both the midterm long-answer problem-solving questions (r_S=0.43) and the midterm short-answer concept questions (r_S=0.42), indicating that conceptual understanding as evaluated by the high-stakes measures is moderately correlated with conceptual understanding as evaluated by the low-stakes DCI. The correlation between the midterm short-answer concept questions and the traditional long-answer problem-solving questions (r_S=0.68) confirms the second hypothesis that student performance on the concept questions is correlated with performance on problem-solving questions. These findings confirm research demonstrating significant relationships between conceptual understanding and problem-solving [391–393].

To address the third hypothesis and further evaluate the relationships between the low-stakes evaluation (DCI) and the high-stakes evaluation (Exam Long and Exam Short), stepwise linear regression with backwards elimination was conducted. The final model showed the low-stakes evaluation is related most strongly with only

the long-answer high-stakes evaluation ($F(1,56)=8.35$, $p<0.01$, $R^2=0.13$) such that scores on the long-answer problem-solving questions positively predict scores on the low-stakes evaluation ($\beta=0.38$, $p<0.01$). In other words, removing the short-answer concept question scores as a predictor variable did not significantly change the fit of the model ($F(1,54)=1.84$, $p=0.18$). However, as was shown in Table 9.2 above, the short-answer concept question scores are in fact strongly correlated with the long-answer problem-solving question scores. The regression analysis was therefore re-conducted with the long-answer scores removed as a predictor variable to verify the relationship between the short-answer concept questions and the low-stakes evaluation. The model revealed the low-stakes evaluation is also strongly related with the short-answer high stakes evaluation ($F(1,56)=7.81$, $p<0.01$, $R^2=0.12$) such that the scores on the short-answer concept questions positively predicts scores on the low-stakes evaluation ($\beta=0.37$, $p<0.01$). As expected, this relationship is nominally weaker than that between the low-stakes evaluation and the long-answer problem-solving question scores.

Students performance on the low-stakes evaluation (post) was predicted by their performance on the long-answer problem-solving questions on the midterm evaluation. Furthermore, their low-stakes evaluation (post) was also predicted by their performance on the short-answer concept questions on the midterm evaluation. Coupled with the moderate Spearman correlation coefficient, this confirms the third hypothesis that students performance on the low-stakes evaluation is predicted by the conceptual understanding demonstrated on both high-stakes evaluations.

Conclusion

Overall, the study described in this section supports the finding reported by Jorion et al. [43] that student performance on the DCI in a low-stakes setting is representative of their understanding of the dynamics concepts covered by the CI [394]. As such, the results from the DCI appears to be an authentic representation of overall student conceptual understanding of the dynamics concepts addressed by the DCI.

9.2.2 Study 2: DCI Instrument Evaluation

Study Background and Objective

While the first study offers evidence of the DCI's ability to quantify student conceptual understanding, some concerns about the reliability and validity of the DCI remain, which will be addressed in this second study. This is achieved using the analytical framework for evaluating CIs outlined by Jorion et al. [43, 385]. The authors centered the psychometric-inspired framework on the validity and reliability of potential interpretations and uses of CI scores and not necessarily of the CI itself [43]. This means the results do not indicate whether the CI is or is not valid; rather, the CI is evaluated based on the level of its plausibility and appropriateness for those potential interpretations and uses [395].

The three claims underlying CIs for educational research are that students' CI scores indicate: 1) their overall understanding of all concepts identified in the CI, 2) their understanding of specific concepts, and 3) their propensity for misconceptions (or student errors) [43]. Individual items (i.e., questions) are evaluated using classical test theory (CTT) [385] and item response theory (IRT) [43] to determine if any items have poor properties that would warrant their removal from the CI. After removing problematic items, the framework recommends conducting factor analyses (exploratory and confirmatory) as well as defining a diagnostic classification model to ensure that the CI's structure is consistent with what the developer intended [43]. As Jorion et al. [43] show, this procedure identifies several items on the DCI as exhibiting very poor properties, which means the inclusion of these items is introducing unnecessary noise into the measurement of overall conceptual understanding. Furthermore, Jorion's exploratory factor analysis indicated very poor structural qualities, which indicates that the 29 items on the DCI do not map well onto the 14 concepts (listed in Appendix B). Thus, the authors found support for the first claim that the DCI can be used to indicate student overall understanding of the concepts in the DCI, but not for the other two claims [43]. As such, the objective of this second study is to conduct CTT and IRT analyses in a local context (i.e. with data collected in ME240 classrooms) to determine if any items warrant removal for subsequent statis-

tical analyses evaluating the effects of the engaged learning IMU intervention (Study 3).

Methods

The DCI items are evaluated using data collected during the pilot study described in Section 8.2 [396]. This data set includes 329 students who elected to participate in the study in either the control or intervention group (note Vernon et al. [396] found no statistically significant differences for overall DCI scores between the groups). Items were evaluated using the two approaches (CTT and IRT) thoroughly outlined by Jorion et al. [43, 385] and briefly described next. These two approaches are similar and can be considered complementary in determining which (if any) items are potentially problematic in subsequent analysis of CI overall scores.

Classical test theory (CTT) is regarded as the "true score theory" and assumes that each measured score is actually the true score plus some error. Cronbach's alpha provides an estimate of CI score reliability, meaning a student's score would be nearly the same if the CI were administered multiple times to the same student. Acceptable values for Cronbach's alpha are typically greater than 0.7, though measures between 0.8 and 0.9 are desirable [397]. To evaluate the effectiveness of a CI's ability to quantify performance, a confidence interval is constructed to represent a range of scores that are not distinguishably different [398]. The greater the reliability (Cronbach's alpha), the smaller the standard error of measurement is and the greater the ability of the CI to measure the students' true scores (a maximum value for reliability would yield a standard error of measurement of 0).

Three additional analyses for CTT investigate the quality of individual items. The first is an assessment of item difficulty, or the proportion of students who answered an item correctly. Ideally, items are not too easy (item difficulties greater than 0.8) or too difficult (item difficulties less than 0.2). The second is an assessment of item discrimination, which is the correlation between how a student answered an item (correctly or incorrectly) and their overall score. A larger value means the item can discriminate between students with high and low total scores, and acceptable values are greater than 0.2. Finally, Cronbach's alpha-with-item-deleted indicates how well

an item fits with the rest of the items and is consistent with the total score. If Cronbach's alpha-with-item-deleted is greater than overall Cronbach's alpha, then the CI is more reliable without the item. Nunnally and Bernstein [397] and Clark and Watson [399] recommended CTT analyses be conducted on samples of at least 300 observations. Thus, the sample of 329 is sufficient.

Item response theory (IRT) considers the relationship between a student's performance on an item and the student's overall performance on the CI. This relationship between item and overall performance is assumed to be a probabilistic logistic model that relates a student's ability (as a value on a continuous scale) to the probability of that student answering that item correctly [400]. Three common model types (1PL, 2PL, and 3PL) of increasing complexity are traditionally used. As Jorion et al. summarize, "*The 1PL has an item difficulty parameter for each item, the 2PL model has an additional discrimination parameter for each item, and the 3PL model has these two parameters plus a guessing parameter to account for the forced choice aspect of multiple-choice questions*" [43, p. 486]. The models are compared with a likelihood ratio test to determine which fits are significantly different from one another with the expectation that including more parameters in the model should significantly improve the fit [401–403]. The reliability of an IRT analysis is a function of both the instrument length and sample size [404]. A simulation study determined for a 30 item instrument that the recommended minimum sample sizes for the 1PL, 2PL, and 3PL models were 150, 250, and 350, respectively [405]. Thus, since the DCI has 29 items, the sample of 329 is adequate. The IRT analysis was conducted using the R package "ltm" [406].

Results and Discussion

The mean (standard deviation) of correct responses for the (end of term) DCI for the 329 students included in the pilot study was 13.8 (4.6), or 47.7 %. The DCI as a whole had adequate reliability (α=0.74), which is also the same as what was reported by Jorion et al. [43]. The standard error of estimation for the sample was 2.02, which is also the same as what was reported by Jorion et al. [43]. For a given student with a score of 14, the 68% confidence interval is 11.94 to 15.97,

which means students with scores between 12 and 16 cannot be inferred with 68% confidence to have different true scores. Figure 9.1 illustrates two of the analyses from CTT, namely item difficulty and item discrimination. Items that are potentially problematic are labelled by their question number. Note that Q1, Q5, Q13, Q14, and Q29 have problematic item difficulty while Q5, Q10, and Q19 have problematic item discrimination.

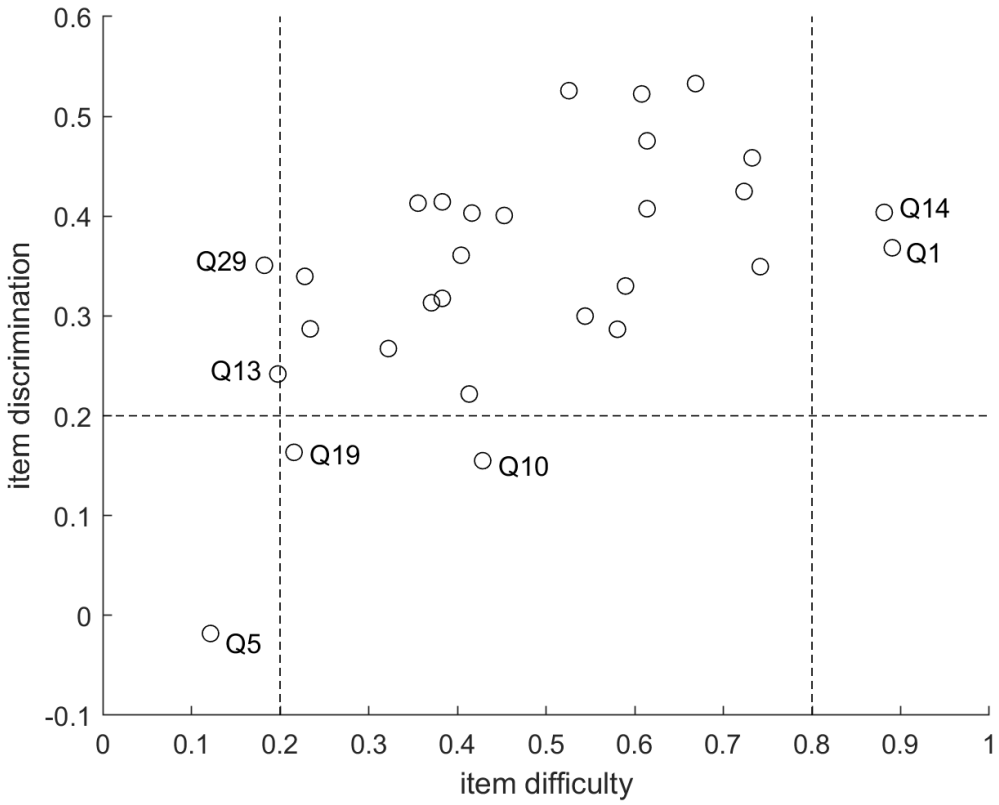


Figure 9.1: A scatter plot of item difficulty and discrimination values for the DCI. The recommended minimum and maximum values are denoted by the dotted lines. Seven items did not meet the recommended values.

Figure 9.2 illustrates the results from the Cronbach's alpha-with-item-deleted analysis. Four items (Q5, Q10, Q19, Q28) produced Cronbach's alpha-with-item-deleted greater than the overall Cronbach's alpha (0.74).

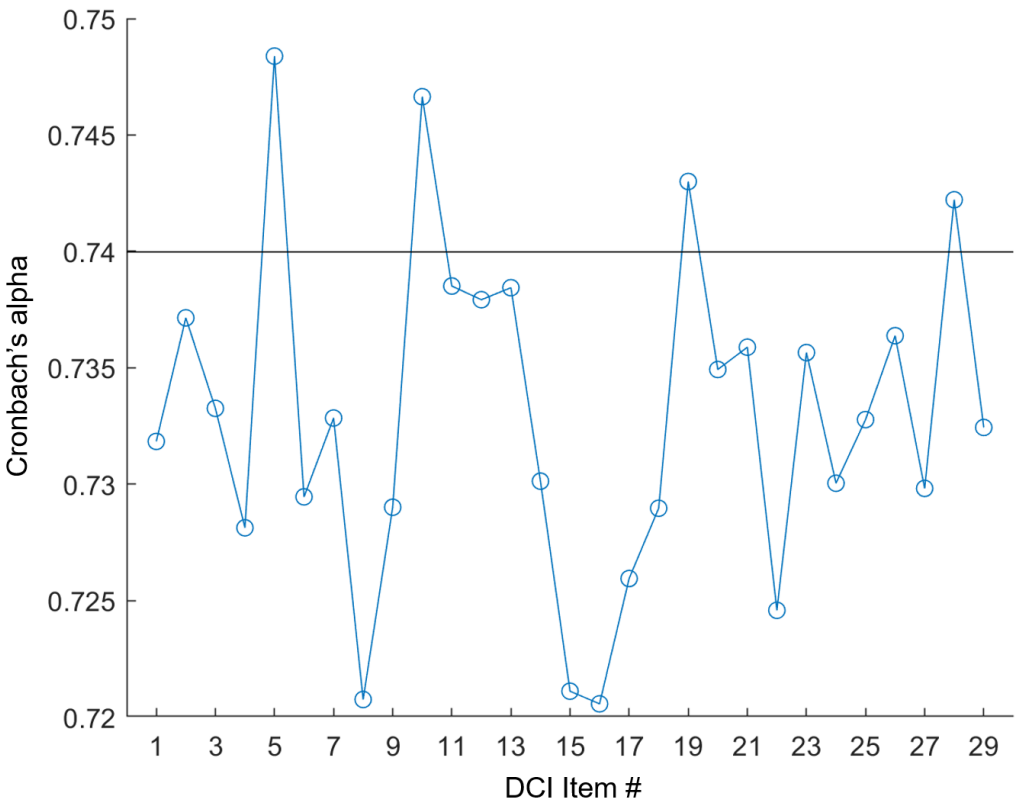


Figure 9.2: A plot of the Cronbach's alpha-with-item-deleted values for each of the DCI items. The solid black horizontal line indicates the overall Cronbach's alpha for the instrument. Four items did not meet the criteria.

Three different logistical models (1PL, 2PL and 3PL) were fit to the data, and likelihood ratio tests were conducted to evaluate which had the best fit. As a reminder, the 1PL model is parameterized by item difficulty, the 2PL model is parameterized by item difficulty and item discrimination, and the 3PL model is parameterized by these two quantities plus a guessing parameter to account for the forced choice aspect of multiple-choice questions. The three models are nested in the sense that the 1PL model is contained within the 2PL and 3PL models and the 2PL model is contained within the 3PL model. The first likelihood ratio test assessed whether the 2PL model was a statistically significantly better fit as compared to the 1PL

model, which the results indicate is the case ($p<0.001$). The second likelihood ratio test assessing whether the 3PL model was a better fit than the 2PL model did not determine it was in fact a better fit ($p=0.17$). Thus, it can be inferred that guessing did not significantly contribute to how students performed on the DCI. The item response curves are shown in Figs. 9.3, 9.4, and 9.5.

Note that Q5, Q10, Q13, Q19, and Q28 are highlighted as potentially troublesome according to this analysis.

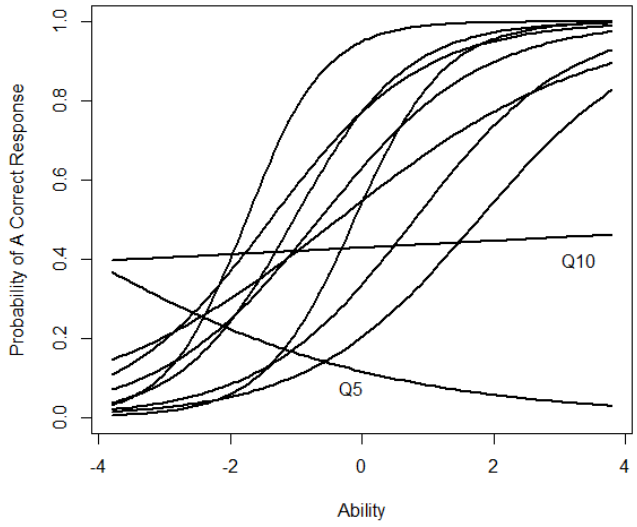


Figure 9.3: Graphs of item response function DCI items 1-10. The shapes for items Q5 and Q10 indicate that these items do not differentiate well between students of high and low ability.

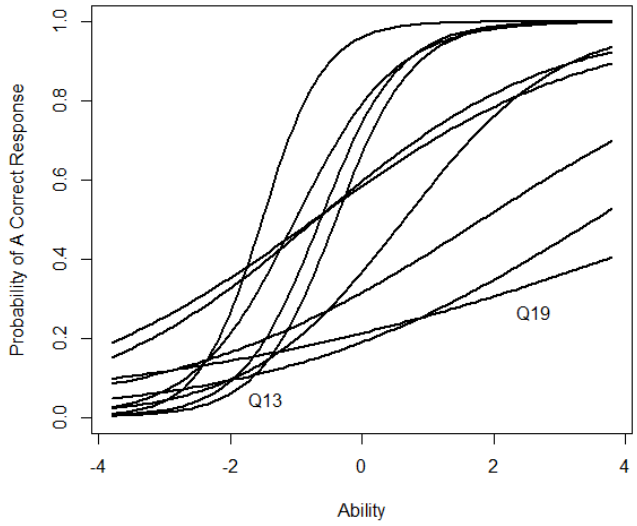


Figure 9.4: Graphs of item response function DCI items 11-20. The shapes for items Q13 and Q19 indicate that these items do not differentiate well between students of high and low ability.

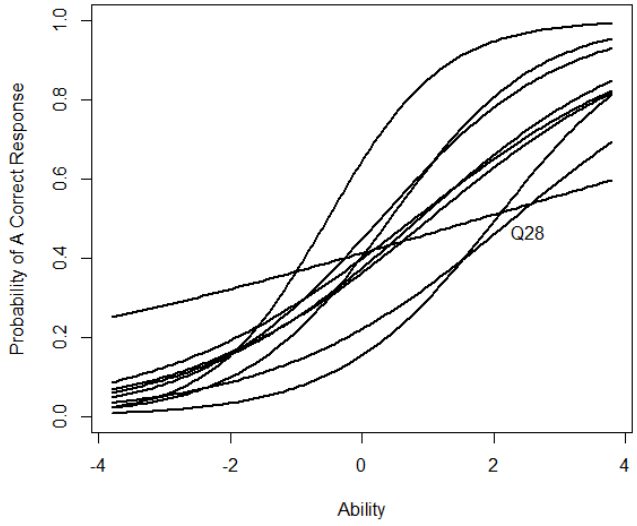


Figure 9.5: Graphs of item response function DCI items 21-29. The shape for item Q28 indicate that this item does not differentiate well between students of high and low ability.

Conclusions

To summarize, item difficulties ranged from 0.12 to 0.89 (Fig. 9.1), and five items had difficulty values outside the recommended range. Item discrimination values ranged from -0.02 to 0.53 (Fig. 9.1), and three items did not exhibit the ability to adequately discriminate between high and low performing students. Four items had Cronbach's alpha-with-item-deleted that were greater than the overall Cronbach's alpha (Fig. 9.2). Recalculating the Cronbach's alpha after removing these items nominally improved the reliability of the DCI (α=0.76). The best fitting IRT model was the 2PL model and five items did not fit the model, thus demonstrating poor ability to discriminate between students of high and low ability. Potentially problematic items identified in all of these analyses are catalogued in Table 9.3 below.

Table 9.3: Summary of flagged items from CTT and IRT analyses.

Approach	**Analysis**	**Flagged Items**
CTT		
	Item Difficulty	Q1, Q5, Q13, Q14, Q29
	Item Discrimination	Q5, Q10, Q19
	Cronbach's alpha-with-item-deleted	Q5, Q10, Q19, Q28
IRT		
	Item Response Curves	Q5, Q10, Q13, Q19, Q28

Given these results and those reported by Jorion et al. [43], four items (Q5, Q10, Q13, and Q28) were removed for the analyses for the final study as they failed at least two inclusion criteria. The total scores for the remaining 25 items retained are then considered to be representative of student conceptual understanding of the engineering dynamics concepts.

9.2.3 Study 3: Active Learning IMU Intervention

Study Background and Objective

This final study leverages the findings from the previous two studies to evaluate the effects of the active learning IMU intervention on student conceptual understanding of engineering dynamics. Study 1 provided support that conceptual understanding evaluated in a high-stakes setting is moderately correlated and predictive of conceptual understanding measured by the DCI in a low-stakes setting [394], thereby allowing the DCI scores to act as a proxy for conceptual understanding. The second study evaluated the reliability and validity of the DCI in a local context (i.e. with data collected from students enrolled in ME240) and identified 4 items to exclude from subsequent analyses. Thus, each students' overall correct scores are calculated from the remaining 25 DCI items for the analyses conducted in this study. Thus, the measurements of conceptual understanding made via the DCI are representative of student understanding of general engineering dynamics concepts covered by the instrument, all of which are covered to varying degrees by the ME240 curriculum.

This study's objective is to investigate the intervention levels' effects on conceptual understanding, including a comparison with a Control group (Level 0) who did not engage with any form of the intervention. As a reminder, the three levels of the intervention are: 1) Demonstrations, 2) Prescribed Experiments, and 3) Student Projects. These levels systematically increase engagement with the IMU technology and cognitive engagement. Specifically, students watch instructor-led demonstrations as passive observers in Level 1, they actively conduct prescribed experiments in Level 2, and they construct their own experiments to study specific concepts in Level 3. According to Chi's ICAP framework [38, 39], higher levels of cognitive engagement hypothetically result in higher gains in conceptual understanding.

Methods

It is common in educational research to evaluate the effects of an intervention on conceptual understanding through learning gains, which have been defined several different ways [407]. Raw gains are simply defined as the difference in scores on an assessment before (Pre) and after (Post) an intervention. Normalized gains are

defined as this raw gain normalized by the amount of room for improvement that exists (i.e. the difference between the Pre-score and a perfect score). In some studies (e.g. Hake [258]), the normalized gains are calculated with the averaged Pre- and Post-scores for an entire classroom. Others extended this definition by calculating normalized gains on an individual student basis (e.g. Marx and Cummings [408]).

However, Theobald and Freeman [407] make a compelling case for using regression analyses to investigate the effects of an intervention. Alone, gains (raw or normalized, averaged or individual) lack the ability to distinguish if learning gains are the result of an intervention or differences in student characteristics. With regression, differences in conceptual understanding at the beginning of the semester can be accounted for by treating the Pre-score as a predictor variable. Regression analyses can also be used to test whether differences in gender, ethnicity, or other factors explain variation in the Post-scores. Finally, regression models can include interaction terms to investigate whether the learning intervention disproportionately impacts different students. Therefore, in this study, multiple linear regression analyses are conducted to evaluate the effects of the active learning IMU intervention on student conceptual understanding. Table 9.4 below specifies each of the outcome and predictor variables as well as the different groups for the categorical variables.

Table 9.4: Summary of multiple linear regression variables. Post and Pre are continuous variables whereas the rest are categorical variables.

Type	**Variable**	**Categories**
Outcome		
	Post	
Predictor		
	Pre	
	Intervention	Level 0, Level 1, Level 2, Level 3
	Gender	Male, Female
	Ethnicity	White, Asian, URM
	Section Type	Large, Small
	Major	ME, AERO, NAME, Elective

Referring to the student demographics reported in Table 8.2, White students are those who identified themselves as not of Hispanic origin and Asian students are those who specifically identified themselves as Asian. Students in the URM (underrepresented minorities) category are those who identified as Black/African American, Hispanic/Latino, American Indian/Alaskan Native, Native Hawaiian/Pacific Islander, or two or more. The small section describes sections with 20 or fewer students whereas large section are those with far more than 20 students. Students who were required to take ME240 as part of their major were in Mechanical Engineering (ME), Aerospace Engineering (AERO), or Naval Architecture and Marine Engineering (NAME). Students who took the class as a technical elective (i.e. did not declare one of those three majors) were broadly categorized as an "Elective" group.

Using the variables described in Table 9.4, the final multiple linear regression model used is

$$Post \sim 1 + Intervention^{*}(Pre + Gender + Ethnicity) + SectionType + Major \quad (9.1)$$

where "1" is a potentially significant intercept and the asterisk indicates that interactions between the intervention and each of the predictor variables in the parentheses are included in the model. It is assumed that each student's response is independent of other students' responses. Also, for the results to follow, normality and heteroscedasticity assumptions are confirmed for the residuals for the model.

Results and Discussion

Table 9.5 contains the full results from the multiple linear regression analysis conducted using the formula described in Eq. 9.1. The model was significant ($F(23, 966) = 52.16$, $p < 0.001$), indicating that it is a significantly better fit than an intercept-only (average-only) model. Overall, the model describes 55% of the variation in the DCI post scores ($R^2 = 0.55$). In the table, CI:LB and CI:UB denote the lower and upper bounds of the confidence intervals for the estimated coefficient. The options in parentheses are the baseline (comparison) categories.

Table 9.5: Summary of (unstandardized) multiple linear regression results.

Variable	Coefficient (β)	CI:LB	CI:UB	t-value	p-value
(Intercept)	16.4	9.5	23.3	4.7	$< 0.001^{\ddagger}$
Pre	0.9	0.7	1.0	12.2	$< 0.001^{\ddagger}$
Intervention (Level 0)					
Level 1	-0.7	-9.0	7.6	-0.2	0.87
Level 2	1.9	-6.4	10.2	0.4	0.66
Level 3	2.9	-6.0	11.9	0.6	0.52
Gender (Male)					
Female	0.5	-5.9	6.8	0.1	0.88
Ethnicity (White)					
Asian	-3.6	-8.4	1.1	-1.5	0.13
URM	1.1	-6.2	8.4	0.3	0.77
Major (ME)					
AERO	-0.9	-2.9	1.0	-0.9	0.35
NAME	1.0	-3.0	4.9	0.5	0.63
Elective	-3.2	-6.8	0.3	-1.8	$0.07^{\bullet}$
Section Type (Large)					
Small	5.0	1.5	8.7	2.8	$< 0.01^{\dagger}$
Interactions					
Level 1*Pre	-0.1	-0.2	0.1	-0.4	0.68
Level 2*Pre	-0.1	-0.2	0.1	-0.8	0.44
Level 3*Pre	-0.1	-0.3	0.1	-1.1	0.27
Level 1*Female	-0.9	-8.2	6.4	-0.2	0.81
Level 2*Female	-4.4	-11.8	3.0	-1.2	0.24
Level 3*Female	-7.2	-14.9	0.5	-1.8	$0.07^{\bullet}$
Level 1*Asian	-0.6	-6.6	5.4	-0.2	0.83
Level 2*Asian	0.4	-5.8	6.6	-0.4	0.91
Level 3*Asian	2.3	-4.2	8.9	0.1	0.48
Level 1*URM	-1.6	-10.2	7.0	-1.4	0.71
Level 2*URM	-6.1	-14.7	2.5	0.7	0.16
Level 3*URM	-4.6	-13.4	4.2	-1.0	0.31

Significant $\alpha = {}^{\bullet}0.1$, ${}^{*}0.05$, ${}^{\dagger}0.01$, and ${}^{\ddagger}0.001$.

The significant intercept and Pre-score coefficient describes on average how conceptual understanding changed over a semester regardless of intervention level or specific student demographics. The intercept means a student who answered no DCI items correctly at the beginning of the semester can answer 16.4% (about 4) items correctly at the end of the semester. The significant Pre-score coefficient (β_{Pre}) means students earned 0.9 Post-score percentage points for every Pre-score percentage point. Figure 9.6 illustrates these two predictor variables relative to the baseline case (no change in conceptual understanding as measured by the DCI).

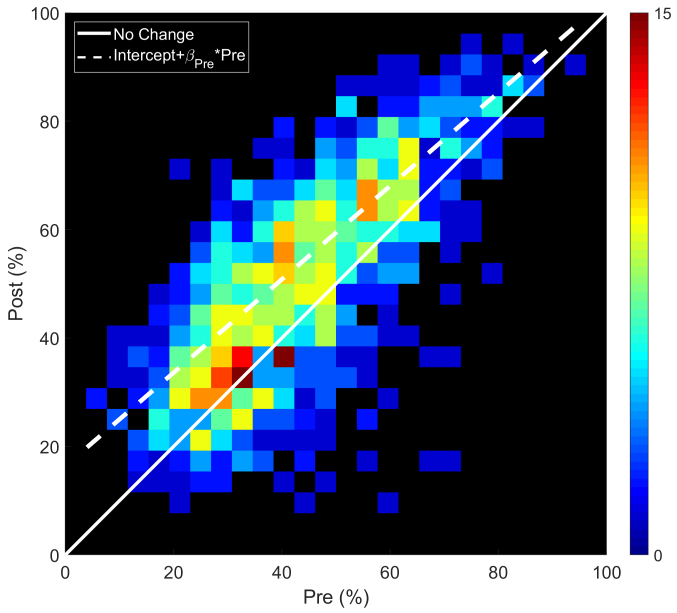


Figure 9.6: The background is a heat map of the joint distribution of Pre- and Post-scores (black areas are void of data). The solid white line denotes the baseline case when the Pre- and Post-scores are equal (no change in conceptual understanding as measured by the DCI). The dashed line illustrates the significant intercept (16.4) and significant slope ($\beta_{Pre} = 0.9$) relating Pre- and Post-scores.

Given the spread of the joint distribution of Pre- and Post-scores, there is a great deal of variation in student performance on the DCI in this study's data set. As is further illustrated by the dashed white line, the majority of the distribution is

concentrated slightly above the baseline case (i.e. the white slide line denoting no change in conceptual understanding). Discussions of the results for the remaining predictor variables in Table 9.5 follow next. While each discussion pertains to a subset of predictor variables, the results presented are with respect to controlling for variations due to the other variables.

Intervention

While Fig. 9.6 illustrates the joint distribution of Pre- and Post-scores for all students, the distributions for each of the intervention levels are informative as well. Figure 9.7 contains split violin plots describing each intervention level's Pre- (grey) and Post-score (gold) distributions. The black boxes at the middle of each violin are box plots corresponding to each distribution with the black horizontal lines indicating the median. Across intervention levels, it is clear from the seemingly equal differences between Pre- and Post-score distribution medians that there is a consistent improvement in conceptual understanding as measured by the DCI.

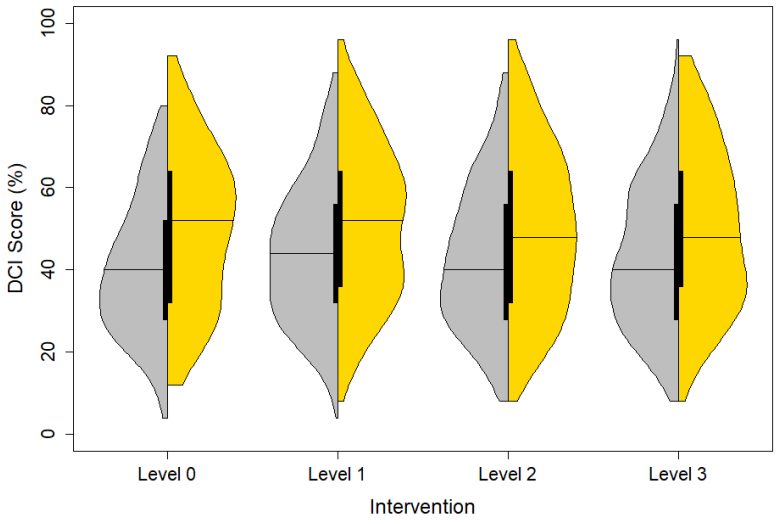


Figure 9.7: Split violin plots of the Pre- and Post-score distributions (grey and gold, respectively) for each of the intervention levels: Level 0 (Control), Level 1 (Demonstrations), Level 2 (Prescribed Experiments), and Level 3 (Student Projects).

However, as is evidenced by the largely overlapping distributions in Fig. 9.7, there were no significant differences in the improvements in conceptual understanding as measured by the DCI across the intervention levels. Figure 9.8 illustrates the slight differences between intervention levels as well as the interactions between the intervention levels and Pre-scores. The nonsignificant interactions mean the interventions had equivalent effects on student conceptual understanding regardless of their conceptual understanding at the beginning of the semester.

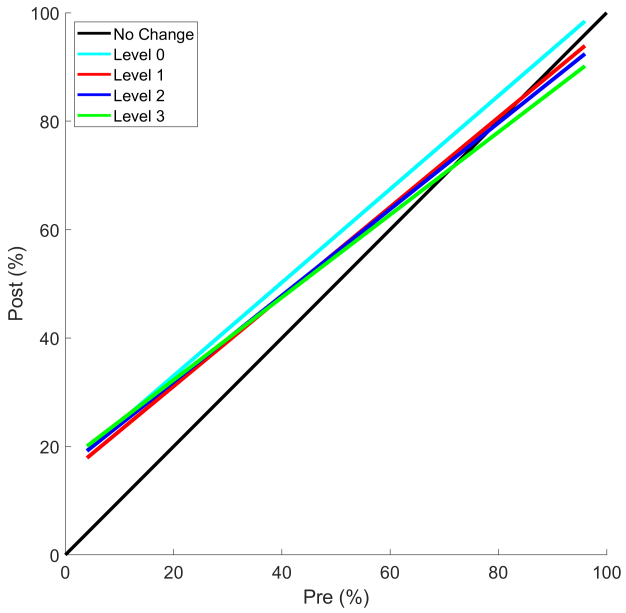


Figure 9.8: The solid black line denotes the baseline case when the Pre- and Post-scores are equal (no change in conceptual understanding). The other lines represent the multiple regression results for each intervention level while controlling for variations in the other predictor variables: Level 0 (Control), Level 1 (Demonstrations), Level 2 (Prescribed Experiments), and Level 3 (Student Projects).

There are many reasons why no significant differences in conceptual understanding are detected, which Chi's ICAP framework hypothesizes should be the result of increased cognitive engagement [38, 39]. First and foremost, the three levels involving the iNewton IMUs were conducted outside of the classroom and in groups of

two or three. Given that there is no regulation for how the students completed the assignments, it is possible that not all of the students fully engaged with the activities as they were intended. As Chi [38] notes, it matters how the students actually engage with the activities, not just how the instructors intended the students to be engaged. This information is not available with the quantitative data collected for this study, but it could be procured in the future using qualitative methods.

Next, there is also the fact that this is a very modest intervention as compared to the rest of the course. Each of the intervention levels involving the iNewton IMUs effectively consisted of two administrations of the intervention among eleven other homework assignments, two midterm examinations, and a final examination, and in addition to the other classes students take concurrently with ME240. As was noted by Hake [258], activities that facilitate a "small amount" of student engagement will not produce the same learning gains as those with a lot of engagement. It is possible that the frequency and intensity of the interventions as they have currently been implemented is not "enough" to yield significant learning gains.

It is also possible that the DCI is too blunt an instrument to measure the learning gains in conceptual understanding resulting from the active learning IMU intervention. Jorion et al. [43] and Study 2 of this chapter both determined that the DCI can be used to assess general conceptual understanding of the engineering dynamics concepts covered by the instrument, not all of which are addressed by the active learning IMU intervention. The types of learning achieved by the students might also not be measured by the DCI as well. For example, in the medical education literature, the effects of problem-based learning (PBL) have been largely measured through traditional modes of assessment (e.g. examinations testing knowledge acquisition) producing roughly 50 years of conflicting results (see, for example, the review and discussion offered by Neville [409]). However, Zahid et al. [410] recently showed that PBL-taught and traditionally-taught students performed the same on questions that test recall of facts or basic comprehension, but PBL-taught students significantly outperform their traditionally-taught peers on questions that test the depth, integration, analysis, and application of knowledge in clinical situations as well as on Objective Structured Clinical Examinations (assessments of clinical skills and pro-

fessional competency). Similarly, it could be the case that what the students are actually learning during the different levels of the active learning IMU intervention are simply not captured by the DCI.

Finally, to gain a sense for how students historically perform on the DCI from past studies, the scores reported in the literature were consolidated and reported in Fig. 9.9. Using all DCI items, the Pre- and Post-scores for the students in this

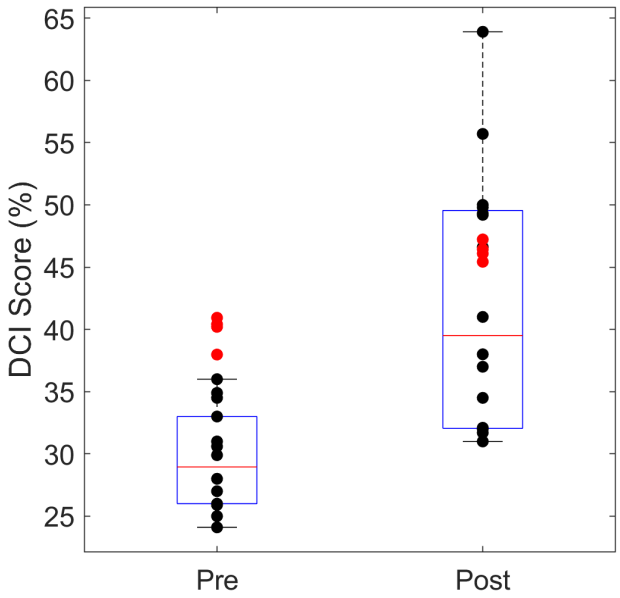


Figure 9.9: The box plots (and solid black dots) are generated from results reported in the literature for approximately 3,930 students involving the DCI administered as either a Pre- and/or Post-evaluation [42, 298, 385, 387–390, 411–413]. The solid red dots are the averaged Pre- and Post-scores for each of the intervention levels in this study (calculated from all 29 questions for comparison).

study's dataset were 40% and 46%, respectively, yielding a 6% raw difference or the equivalent of just 2 more questions answered correctly. The Pre- and Post-score averages for the data consolidated from the literature were 29% and 42%, respectively, yielding a 13% raw difference or 4 more questions answered correctly. Students in this study scored higher on the Pre-evaluation than the reported data and scored on

the higher end of the interquartile range of the reported Post-evaluation data.

Combining this study's data with the data collected from the literature ($N \approx 3,930$), the average Pre- and Post-scores is 32% and 43%, yielding an 11% raw gain (3 questions) or a normalized gain of 0.16. For comparison, the Force Concept Inventory developed by Hestenes et al. [376] has been administered in physics classrooms around the nation, and Hake [258] collected a large subset of this data to assess the effects of interactive engagement in classrooms. That data set ($N = 6,542$) yielded an average normalized gain (regardless of teaching pedagogy) of 0.35, more than double that measured by the DCI. This evidence is offered in support of the conclusion that one of the reasons for the limited impact of the active learning IMU intervention on conceptual understanding could be due to the instrument's inability to measure conceptual understanding on a finer scale.

Gender

Next, Pre- and Post-score distributions by gender are illustrated in Fig. 9.10, and the results from the multiple regression by gender are plotted in Fig. 9.11. Despite

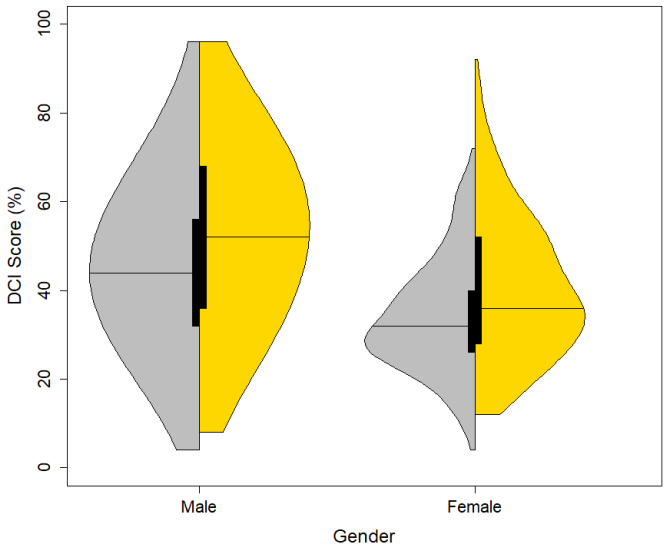


Figure 9.10: Split violin plots of the distributions for each gender. (Pre- and Post-score distributions are grey and gold, respectively.)

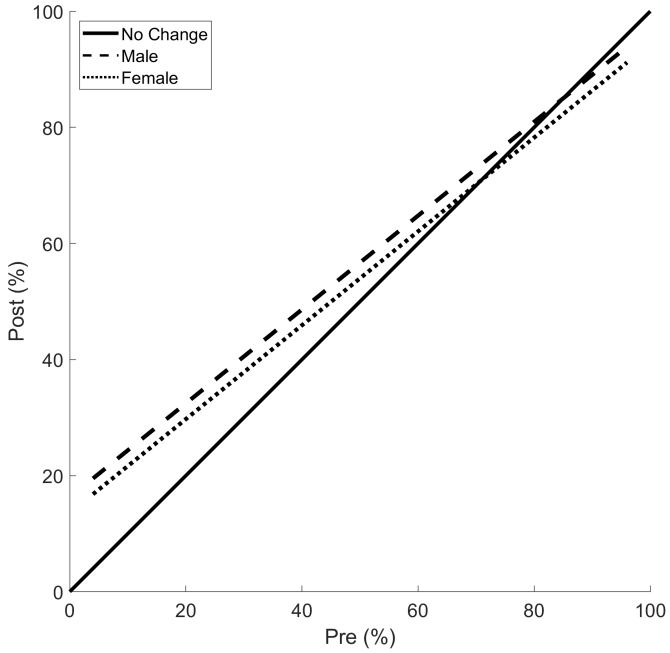


Figure 9.11: The solid black line denotes the baseline case and the other two lines are for each gender.

the fact that female students appear to consistently score lower than their male counterparts (i.e. the median Pre- and Post-scores are lower for female students), the differences are not statistically significant. This nonsignificant finding can partially be explained by the large variation in Fig. 9.10 and the unequal sample sizes of male students ($N = 797$) versus female students ($N = 227$) in the study.

With respect to the relationship between gender and the intervention, the results from the regression analysis indicate the interactions are not statistically significant at $\alpha = 0.05$. However, the results illustrated in Fig. 9.12 show there could potentially be a meaningful interaction between the intervention and female students, particularly for Level 3 (Student Projects). Although the p-value (0.07) does not indicate significance, the magnitude of the estimated coefficient (-7.2) is larger than every other estimated coefficient with the exception of the intercept (16.4). The nonsignificance can at least be partially explained by unequal sample sizes between male

and female students as well as the large variations in Figs. 9.7 and 9.10. Specifically, students in Level 0, Level 1, Level 2, and Level 3 were 14%, 24%, 23%, and 22% female, respectively.

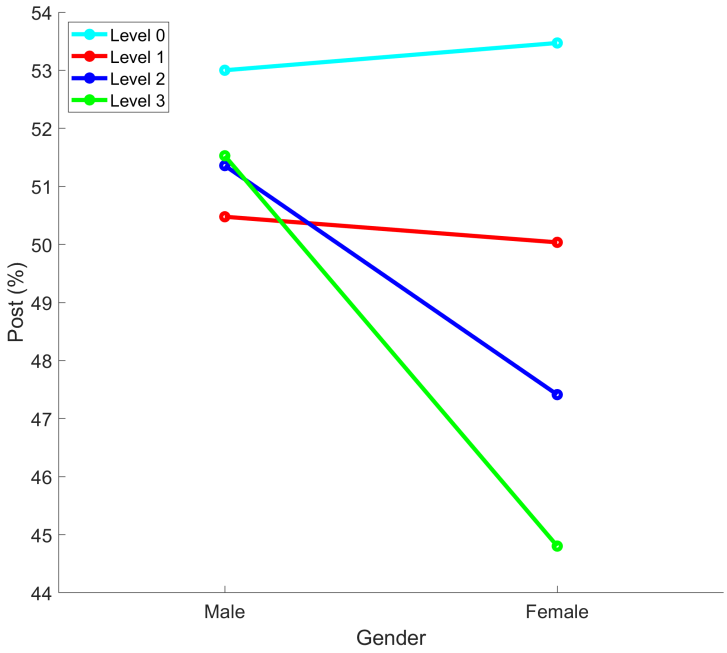


Figure 9.12: The four lines illustrate by gender each of the intervention levels: Level 0 (Control), Level 1 (Demonstrations), Level 2 (Prescribed Experiments), and Level 3 (Student Projects).

However, the trend of increasing female students' engagement with the iNewton IMUs seemingly associated with poorer conceptual understanding is concerning. The difference between Level 0 and Level 3 is roughly 7% or 2 questions on the DCI, which is substantial given the magnitude of the largest estimated coefficient (the intercept) is 16.4% or 4 questions. In Level 2 (Prescribed Experiments) and Level 3 (Student Projects) of the intervention, students worked in self-selected groups. During office hours specifically for the iNewton assignments, it was observed that many groups delegated work to expedite the process. It is possible that the female students were more frequently responsible for secretarial roles (e.g. organizing and writing up the

assignment) and otherwise not as fully engaged with the learning activity (see, for example, Cech [414] or Smith and Gayles [415]). Figure 9.13 shows the composition of Student Project groups by gender for which only 6 of the 116 groups were comprised solely of female students.

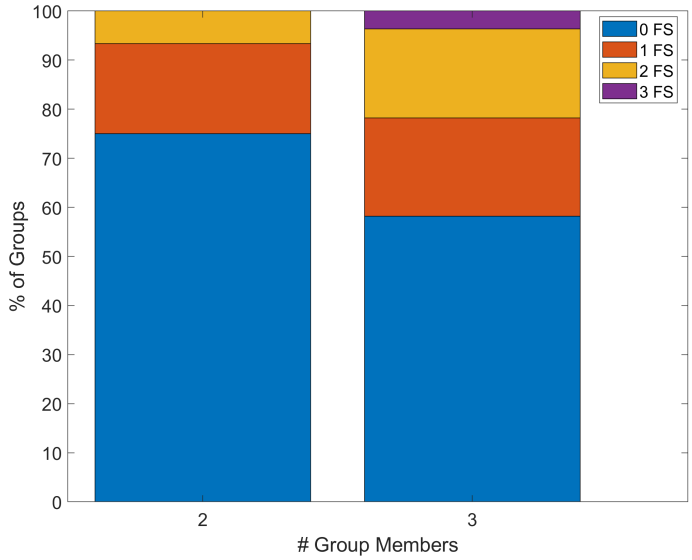


Figure 9.13: Student Project group composition by gender. Each color corresponds to the number of female students (FS) in the group. In total, there were 61 groups of 2 and 55 groups of 3.

In a 2013 study conducted in the University of Michigan's College of Engineering, Meadows and Sekaquaptewa found women often adopt stereotypical roles in group work without perceiving pressure to do so, which the authors explained "*that in adopting stereotypical behaviors, women are taking a specific action to increase their sense of belonging on the team to align with team members biased expectations*" [416, p. 12]. Hirshfield [417] further elucidates this finding by showing undergraduate students may be recognized as gendered stereotypes. For a discussion of implicit bias and female engineering students' socialization in group work, see Seron et al. [418]. However, this speculation can only be answered by interviewing the female students who participated in Level 3 of the intervention.

Ethnicity

Next, the Pre- and Post-score distributions by ethnicity are illustrated in Fig. 9.14. As expected from the distributions illustrated in Figs. 9.7 and 9.10, there is an increase in scores from beginning to end of the semester (evidenced by the median lines). However, there are also large variations in scores across all three groups of students.

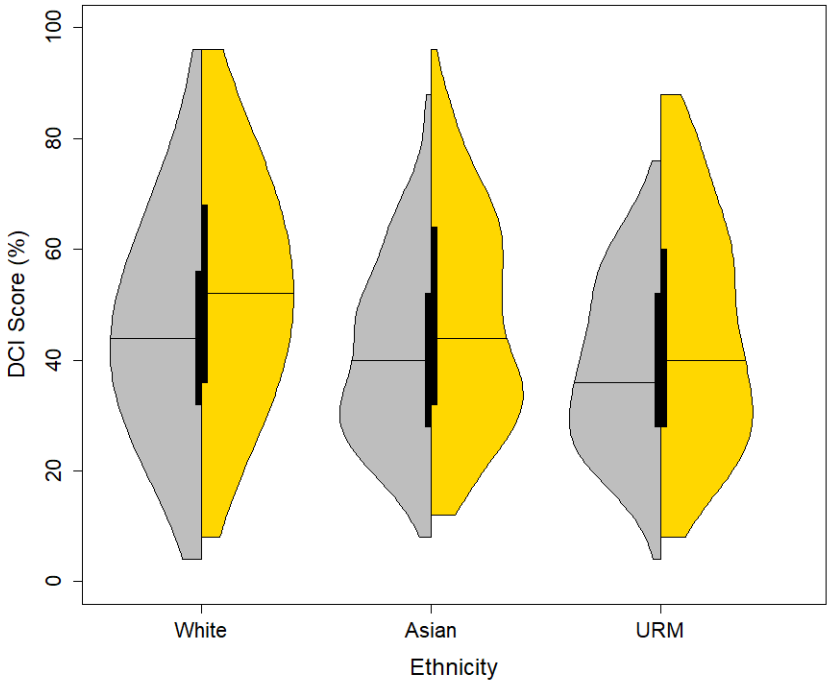


Figure 9.14: Split violin plots of the distributions for each ethnicity. (Pre- and Post-score distributions are grey and gold, respectively.)

The results from the multiple regression by ethnicity are plotted in Fig. 9.15. In both the distributions in Fig. 9.14 and the regression results in Fig. 9.15, White students appear to consistently score higher than Asian or URM students. However, the results from the multiple regression analysis reveal these differences are not statistically significant.

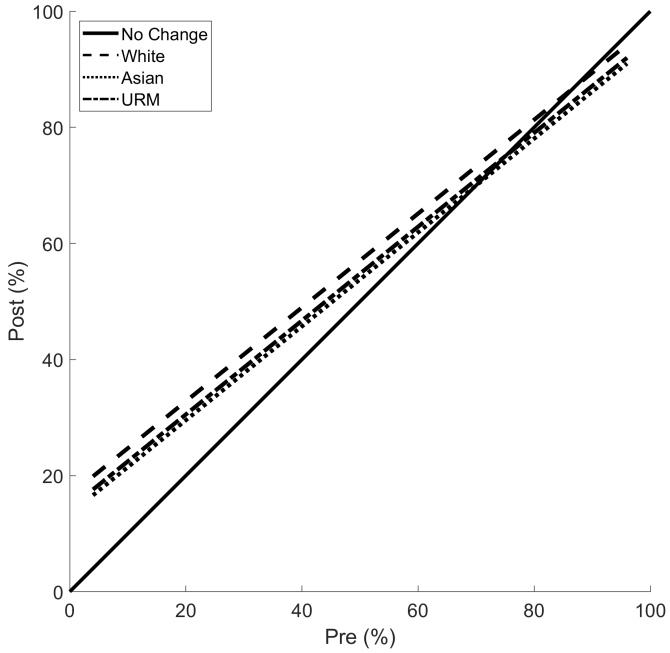


Figure 9.15: The solid black line denotes the baseline case and the other lines are for each ethnicity.

The relationship between ethnicity and the intervention is illustrated in Fig. 9.16, though the results from the multiple regression analysis reveal none of these interactions are statistically significant. Like with gender, these nonsignificant findings can be partially explained by the large variations in the distributions in Fig. 9.14. The URM group in particular is a very heterogeneous collection of students who are otherwise lumped together largely for the sake of statistical power. It is likely that the students in this group have considerably different experiences in preparing for an engineering curriculum, which could partially account for why the distributions for URM students exhibit such large variations in Fig. 9.14. In tandem with the large variation in DCI scores is unequal sample sizes of White students ($N = 657$) as compared to Asian ($N = 203$) and URM ($N = 130$) students. In particular, 15 (10%), 40 (11%), 41 (13%), and 34 (15%) of the students in the Level 0, Level 1, Level 2, and Level 3 intervention groups, respectively, identified as members of

a URM. Large variations in small samples (relative to White students) yields large confidence intervals (i.e., less precise estimates of the true value of the groups' DCI scores), which makes statistical significant difficult to achieve.

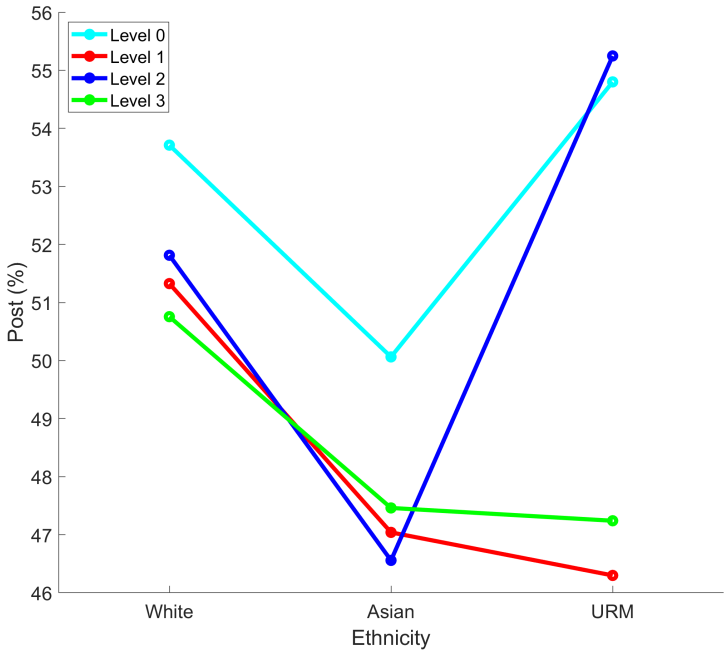


Figure 9.16: The four lines illustrate by ethnicity each of the intervention levels: Level 0 (Control), Level 1 (Demonstrations), Level 2 (Prescribed Experiments), and Level 3 (Student Projects).

Major

Next, Pre- and Post-score distributions by major are illustrated in Fig. 9.17, and the results from the multiple regression by major are plotted in Fig. 9.18. In general, students who take ME240 as an elective underperform on the DCI as compared to their peers who are required to take the course as a part of their major's curriculum. However, this estimated coefficient (-3.2) describing the difference between Mechanical Engineering (baseline) and Elective students is marginally significant (p=0.07).

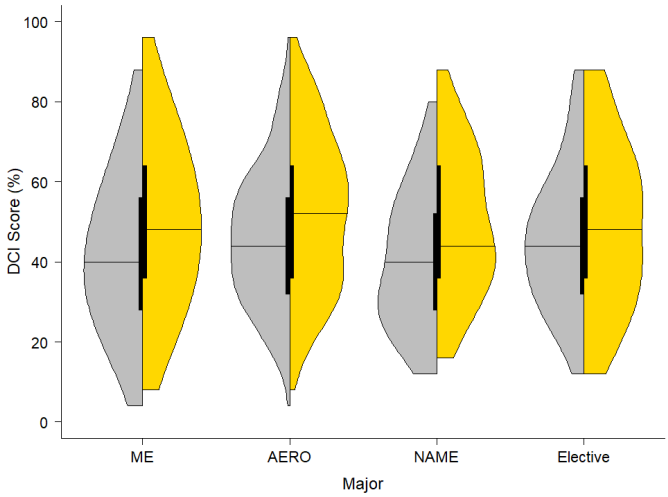


Figure 9.17: Split violin plots of the distributions for each major. (Pre- and Post-score distributions are grey and gold, respectively.)

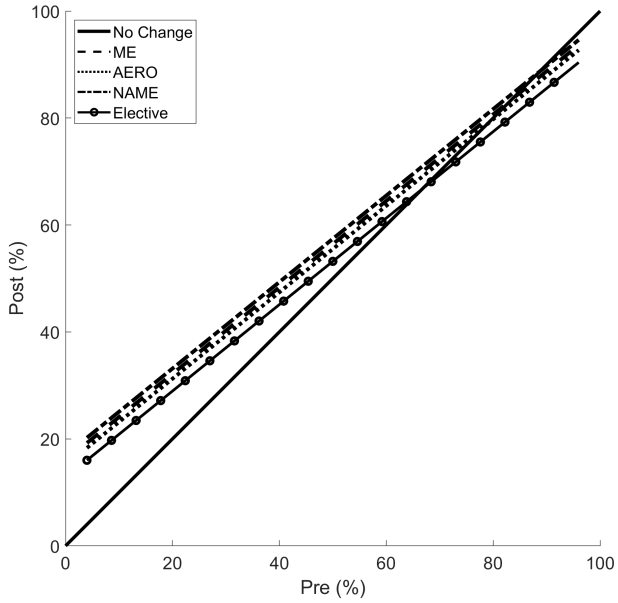


Figure 9.18: The solid black line denotes the baseline case and the other lines are for each major.

Section Type

The different section sizes are the result of a small section initiative at the University of Michigan, which began with the Provost's Office in the form of a request for proposal from major units, including the College of Engineering. The college responded by identifying via department input potential "gateway" courses where students who frequently appear to struggle may benefit from offerings of small sections (i.e., enrollments of 20 students or less). The Mechanical Engineering department identified their sophomore engineering science courses (including ME240) as those gateway courses, and began offering small sections for these courses intermittently starting in the Fall 2016 semester (i.e., the same semester this current study began).

The Pre- and Post-score distributions by section type are illustrated in Fig. 9.19. Referring specifically to the Post-score distributions' medians in Fig. 9.19, students enrolled in Small sections appear to score higher during the end of semester evaluation as compared to their peers enrolled in Large sections.

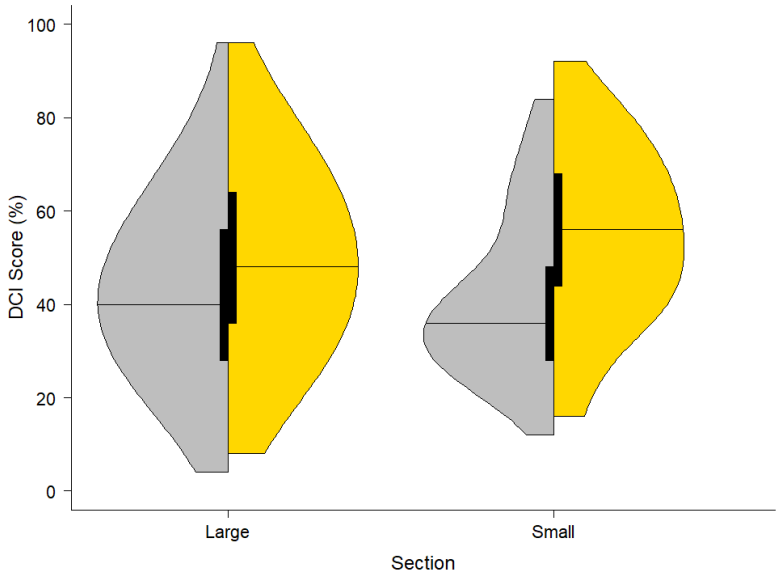


Figure 9.19: Split violin plots of the distributions for each section type. (Pre- and Post-score distributions are grey and gold, respectively.)

The results from the multiple regression are plotted in Fig. 9.20, which also reveal that students in the Small sections outperformed their peers in the Large sections of the class. Educational research across various disciplines also supports the positive effects of small section offerings with implications for students, faculty, and staff. For students, Cheng [419], McDonald [420], and Bradley et al. [421] all found that small section sizes are associated with improved learning gains, critical thinking and problem-solving skills, overall satisfaction, and retention rates. Likewise, Freeman et al. [260] showed active learning has the most significant positive effects on student academic performance in smaller classrooms. Toth and Montagna [422] also identified large class sizes as a barrier of entry for those who enter higher education underprepared, including those who are educationally disadvantaged. For faculty and staff, Cheng [419] and Gillespie et al. [423] described negative effects on overall well-being and morale, student evaluations, and research productivity with increased enrollments.

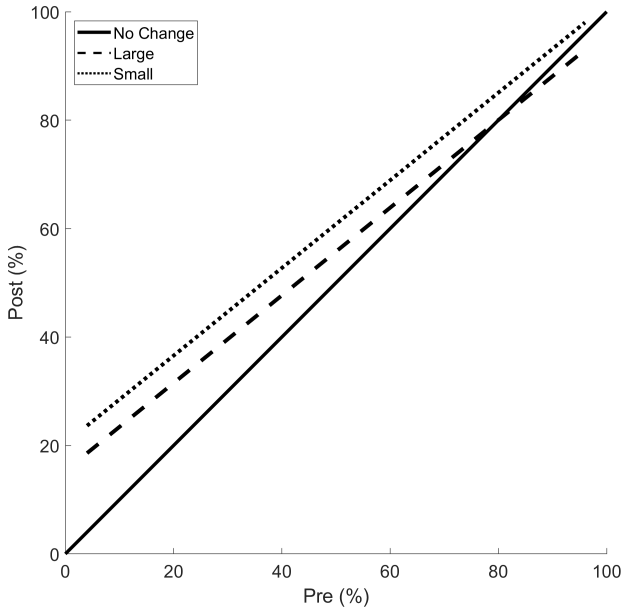


Figure 9.20: The solid black line denotes the baseline case and the other lines are for each section type.

Conclusions

Overall, the results of the statistical analysis investigating the effects of the active learning IMU intervention on conceptual understanding were mixed. Given the nonsignificant estimated coefficients for the intervention levels, the impacts of the intervention levels on conceptual understanding, at least as measured by the DCI, were limited. Future work should consider the interactions between the Student Projects version of the intervention and gender and ethnicity (Figs. 9.12 and 9.16, respectively). Other results not related to the intervention levels were also revealed through the statistical analysis. Differences in student performance by major revealed students who take the course as a technical elective tend to underperform on the DCI as compared to their peers who are required to take ME240 as a part of their major's curriculum. Finally, the results concerning section size confirm the previous findings reported in the literature as to having a significant impact on student learning with smaller sections facilitating greater learning gains.

9.3 Summary and Conclusions

To summarize, this chapter evaluated the efficacy of the DCI administered in a low-stakes setting as a measure of conceptual understanding, investigated the reliability and validity of the DCI as an instrument, and assessed the impacts of three versions of an active learning IMU intervention on conceptual understanding as measured by the DCI. The first study confirmed that conceptual understanding measured in a high-stakes setting (i.e. a midterm evaluation) was moderately correlated and predictive of conceptual understanding measured in a low-stakes setting (i.e. DCI scores at the end of the semester). However, the regression models confirming the predictive relationship only explained roughly 13% of the variation in DCI performance. One reason for this is the DCI covers many more concepts than what was covered in the high-stakes evaluation (a midterm examination). Another reason was revealed in Study 2, which found there were several items on the DCI that were unreliable and subsequently removed for the statistical analysis conducted in Study 3. The nonsignificant regression results in the final study revealed the impacts of the

active learning IMU intervention on conceptual understanding, at least as measured by the DCI, were limited. However, differences in student performance by major and section size were also revealed by the results of the regression analysis.

Finally, an attempt was made to categorize the remaining 25 DCI items by course concepts described by the ME240 curriculum, but the internal reliabilities were unacceptably low (the highest Cronbach's alpha was 0.57). Coupled with the poor structural qualities revealed by an exploratory factor analysis conducted by Jorion et al. [43], the DCI in its current form cannot reliably be used to evaluate changes in conceptual understanding for any subsets of concepts covered by the DCI. Future work should include revisions to the DCI to improve its ability to differentiate levels of conceptual understanding.

CHAPTER 10

iNewton Non-Cognitive Effects

10.1 Background to Studies

Past research identified student self-efficacy (the strength of one's belief in knowledge and skills needed to achieve a task) and conceptual understanding as being related to intention to persist in a STEM major (for a more extensive discussion, refer to Section 7.2 in Chapter 7). The underlying assumption is that improvements in student conceptual understanding and self-efficacy positively interact with one another and with intention to persist. Importantly and relevant to this chapter, Hoit and Ohland [330] showed strategies that foster self-efficacy and intention to persist in the general student population are disproportionately successful for students who are historically underrepresented in engineering, like women and students of color. Albert Bandura stresses the far reaching influences of self-efficacy as "*the courses of action people choose to pursue, how much effort they put forth in given endeavors, how long they will persevere in the face of obstacles and failures, their resilience to adversity, whether their thought patterns are self-hindering or self-aiding, how much stress and depression they experience in coping with taxing environmental demands, and the level of accomplishments they realize*" [45, p. 3].

This chapter examines how the active learning IMU intervention influences student self-efficacy as well as intention to persist and student affect (e.g., how students feel about the IMU intervention), collectively referred to as non-cognitive effects. To

begin, a review and introduction of instruments that measure each of these three effects is presented.

10.1.1 Self-Efficacy

Self-efficacy has been studied at various grain sizes, meaning with respect to the specificity of the skills being assessed. For example, Schwarzer and Jerusalem [424] developed the General Perceived Self-Efficacy Scale to assess a broad sense of personal competence to deal effectively with a variety of stressful situations. Scholz et al. [425] administered this survey to over 19,000 individuals in 25 countries to establish that self-efficacy is a universal construct. Among a population of engineering students, Dunlap [329] used this instrument to track differences in self-efficacy over the course of an academic term after introducing realistic problem-based learning activities into an engineering capstone course. In the study, she concluded that such activities positively influence overall self-efficacy. While these findings (and many others) are pertinent, Bandura recommends "*scales of perceived self-efficacy must be tailored to the particular domain of functioning that is the object of interest*" [426, p. 307-308]. In other words, instruments designed to measure self-efficacy need to be developed for a specific context, and the degree of the specificity will vary with that context.

To that end, a more focused instrument is the Mathematics Self-Efficacy Scale, which was developed by Betz and Hackett [427] to investigate differences in mathematics-related behaviors by gender. While this instrument focuses on a specific domain (mathematics), it is general in the sense that it can be applied to students in domains other than mathematics with little to no alteration. For example, Morán-Soto and Benson [428] used this instrument in a mixed-methods study to determine the relationship between first-year engineering student mathematics self-efficacy and preparation. Another example of an instrument focused on a domain-specific skill is the Engineering Design Self-Efficacy Instrument developed by Carberry et al. [429]. In that study, self-efficacy differed between experienced and novice engineers, correlated positively with student motivation and outcome success expectancy, and correlated negatively with anxiety. Finally, an example of a singularly focused instrument

is found within the Motivated Strategies for Learning Questionnaire developed by Pintrich et al. [430, 431]. Items on this instrument probe student knowledge and skills for a specific class, though the questions are applicable to any course (e.g. "I am confident I can learn the basic concepts taught in this course.") [430]. Fang [313] used this instrument to evaluate the relationship between course-specific self-efficacy and academic performance in an undergraduate dynamics course, finding that self-efficacy was significantly and positively correlated with performance.

Self-efficacy can also have longitudinal variations and thus change over time. In a study investigating the relationship between career interests and general self-efficacy, Nauta et al. [432] found no significant differences in self-efficacy after three or seven months in psychology undergraduate students. The instrument used in that study was the Skills Confidence Inventory, which was originally developed by Betz et al. [433] to evaluate gender differences in self-efficacy. Similarly, no changes in self-efficacy were detected after five months in first year engineering students using an instrument designed by Lent et al. [434] to measure self-efficacy for academic milestones and coping. However, in a follow-up study, Lent et al. [435] did report statistically significant (negative) changes in self-efficacy measured by the same instrument in the same population after 2 years. The Longitudinal Assessment of Engineering Self-Efficacy (LAESE) is an instrument specifically designed to study changes in self-efficacy over time. Most of the studies that employ the LAESE (described in Section 10.1.4) typically administer it annually. For example, Payton et al. [436] contextualized the LAESE while measuring the positive effects of a service learning intervention over the course of an academic year. These studies indicate that changes in self-efficacy could depend not just on time but also other factors not considered in this research (e.g., intensity or frequency of an intervention).

10.1.2 Intention to Persist

The following review is focused on the ways intention to persist has been measured in the literature, after briefly discussing the construct itself. Persistence, retention, and attrition are different but related metrics frequently used (and sometimes in-

terchangeably) in education research and institution assessments. Retention and attrition are usually terms reserved for evaluating an institution's ability to keep students (retention) or not (attrition). However, this can be misleading when one considers, for example, transfer students who stay in the field but move between institutions (e.g. community colleges). Persistence refers to an individual's choice to stay in a specific program, field, college, or institution.

The notion of intention to persist follows from Ajzen's Theory of Planned Behavior [437, 438], which describes intention as preceding subsequent behavior. This was adapted by Bean's model [35] for student attrition in which he operationalized persistence as the intent to leave to find that it, along with grades, were the strongest predictors of student attrition. Subsequently, researchers developed various measures of intention to persist. The most common method employs one or a set of Likert-type items. For example, Hausmann et al. [439] use a single item (e.g. "I intend to complete my degree at <name of institution>") to determine its relationship with sense of belonging and ethnicity. Similarly, Hatch and Garcia [440] employed a single item to investigate the relationship between advising and intention to persist for community college students. The Persistence in Engineering survey developed by Eris et al. [441] also includes a single item to measure intention to persist. In the absence of a well-established measure of intention to persist, Wheeless et al. [442] created a six-item survey to study undergraduate students' intention to continue their education. Robinson [443] developed the Undergraduate Persistence Intentions Measure to measure undergraduate persistence/drop-out intentions across two categories, namely: 1) Persistence at the Institution and 2) Persistence in Higher Education. Davidson et al. [444] developed the College Persistence Questionnaire, which operationalized intention to persist as Institutional Commitment and Degree Commitment. Recently, Pugh et al. [445] altered that questionnaire specifically to measure intention to persist in nursing students. Within the LAESE, Marra et al. [46] included three items related to intention to persist, though the creators did not designate this set of items to be a specific construct measured by the instrument. Importantly, all of these studies successfully employed their respective measures of intention to persist in the context of larger related research question(s).

10.1.3 Student Affect

Student affect is defined as the students' feelings towards or interest in a specific activity, which in this research is the active learning IMU intervention described in Chapter 8. Hidi [446] described interest as determining how one preferentially persists in processing certain types of information, which inevitably leads to improved performance. Hidi and Renninger [447] expanded on this by showing personal characteristics and social contexts can also influence the development of interest when engaging in an activity. In a meta-analysis of more than 150 studies, Schiefele et al. [448] found statistically significant correlations between interest and both academic and laboratory performance ($r_s = 0.31$ and $r_s = 0.27$, respectively). The research of Harackiewicz and Hulleman [449] revealed interest can play a significant role in future choices and career paths. Finally, Betz et al. [433] found significant correlations between self-efficacy and interest for both undergraduate male and female students. Accordingly, student affect is an additional effect that may provide useful information for comparing the active learning IMU intervention levels.

Students often respond positively towards active learning innovations implemented in undergraduate engineering dynamics classrooms. For example, Ferri and Ferri [300] implemented several hands-on experiments facilitated by IMUs, and students reported preferring and learning more from those experiments compared to traditional instructional methods. Bevill and Bevill [375] similarly found students enjoyed collecting and analyzing IMU data outside of a laboratory setting and felt the IMU activity contributed positively to their conceptual understanding of particle dynamics. In a crossover study comparing demonstrations and hands-on experiments, Self and Widmann [450] found no student preference for one active learning mode over the other, but they observed that either mode yielded significantly more learning gains compared to traditional instructional modes. Finally, O'Connor [377] reported predominantly positive feedback from students who designed and conducted experiments of their own imagining using the inertial sensors embedded in their smartphones. Accordingly, student affect provides another means to assess the impact of the active learning IMU intervention and at each level of implementation.

10.1.4 The Longitudinal Assessment of Engineering Self-Efficacy

This research employs the LEASE to measure non-cognitive effects, and a summary of that instrument and some history of its development is provided here.

In 2004, as a part of the NSF-funded Assessing Women in Engineering project, Marra, Moore, Schuurman, and Bogue describe the development and early results of an instrument to "*identify sources of barriers or obstacles in the task of obtaining an engineering degree and ascertain how capable a person feels in those situations*" [362, p. 6]. The instrument, now known as the Longitudinal Assessment of Engineering Self-Efficacy (LAESE), was designed to measure self-efficacy in women who are studying engineering [46], and the authors piloted it to confirm the reliability and validity of the subscales. Since its conception, the LAESE instrument has undergone several revisions, and the latest version has 31 Likert-type items assessing 6 constructs, namely: 1) Engineering Self-Efficacy I, 2) Engineering Self-Efficacy II 3) Engineering Career Success Expectations, 4), Feeling of Inclusion, 5) Coping Self-Efficacy, and 6) Math Outcome Expectations. Engineering Self-Efficacy I focuses on students' self-efficacy in an engineering major in general. Engineering Self-Efficacy II is focused on students' self-efficacy in their engineering curriculum. Engineering Career Success Expectations measures students' beliefs in finding a job they would like. Feeling of Inclusion assesses the degree to which students' feel like they are a part of their engineering community. Coping Self-Efficacy measures students' ability to cope as engineering students. Finally, Math Outcome Expectations evaluates students' perception of the value and utility math in engineering.

Studies employed the LAESE to explore various facets of self-efficacy across different engineering student demographic groups. Marra et al. [46] collected longitudinal LAESE data and provided evidence that women may experience a loss of self-efficacy in the first year of an engineering curriculum. In a follow-up study collecting data from multiple institutions for two years, the authors found somewhat conflicting results in that several constructs increased while others decreased [308]. Specifically, Engineering Self-Efficacy II, Coping Self-Efficacy, and Math Outcome Expectations increased over the year-long study (including 196 women engineering

students), whereas Engineering Self-Efficacy II, Engineering Career Success Expectations, and Feeling of Inclusion decreased. The results did not reveal any significant differences by institution or year standing.

Using an altered version of the LAESE, Concannon and Barrow [451, 452] confirmed the results above in finding no significant differences in engineering self-efficacy from year-one students to year-four students regardless of gender as well as no significant differences between first- to fifth-year women. However, in another study [453], the authors then found womens intentions to persist in undergraduate engineering were more dependent on their ability to achieve higher academic performance than their male counterparts. In other words, women's beliefs in their ability to obtain an 'A' or 'B' in their courses were predictive of their persistence whereas in men they were not. Jordan et al. [454] also found that engineering self-efficacy did not vary across first- and second-year engineering students, but URM students' engineering self-efficacy was significantly lower than that of majority students. Ranalli and Ritzko [455] administered an altered version of the LAESE to evaluate the effects of video gamebased design projects on changes in self-efficacy for first-year engineering students, and found that Engineering Self-Efficacy I declined while Technology Self-Efficacy and Communication Self-Efficacy improved. Finally, Gaikwad and Kulkarni [456] measured changes in self-efficacy for 600 women engineering students in India across 2 years using the LAESE and found significant decreases in Engineering Career Success Expectations and Feelings of Inclusion over time.

10.2 Four Studies Concerning Non-Cognitive Effects

The remainder of this chapter is organized into four studies focused on the non-cognitive effects the active learning IMU intervention has on students in ME240. The first study considers the validity and reliability of a modified version of the LAESE as an instrument measuring self-efficacy (engineering and course-specific), intention to persist, and student affect. The remaining three studies leverage these results to investigate the impact of the active learning IMU intervention on 1) self-efficacy, 2) intention to persist, and 3) student affect, respectively.

10.2.1 Study 1: LAESE Instrument Evaluation

Study Background and Objective

This study considers two types of self-efficacy, namely engineering self-efficacy and course-specific self-efficacy. Because the mastery experience provided by the active learning IMU intervention could contribute to students' perceptions of their skills and abilities, it could impact their self-efficacy in engineering (engineering self-efficacy) as well as in understanding and applying engineering dynamics (course-specific self-efficacy). The LAESE instrument explicitly includes two constructs measuring engineering self-efficacy (Engineering Self-Efficacy I and Engineering Self-Efficacy II) and as well as an item related to course-specific self-efficacy.

Concerns over the validity and reliability of the LAESE instrument have been raised including some relating to the two Engineering Self-Efficacy constructs of interest to this research. For example, Mamaril et al. [457] noted the ambiguous wording of several items that may assess expectancy of success rather than self-efficacy. Concannon and Barrow [452] reported a strong correlation ($r = 0.83$) between Engineering Self-Efficacy I and II; a correlation of 0.85 or greater implies these two constructs are not likely distinct from one another [458]. A recent study by Yoon and Sorby [459] catalogued concerns and documented a thorough analysis of the reliability and validity of the LAESE. Included in that analysis was an exploratory factor analysis and multiple confirmatory factor analyses to determine the best fit model. Yoon and Sorby [459] offered a final revised version which reduced the number of items on the LAESE from 31 to 16 and comprised 4 constructs with acceptable validity and reliability, namely: 1) Engineering Self-Efficacy (6 items), 2) Engineering Career Expectations (3 items), 3) Sense of Belonging (3 items), and 4) Coping Self-Efficacy (4 items). The six items identified in their study as measuring Engineering Self-Efficacy are used in this study. Also, two items were added to the original single item in the LAESE to create a Course-Specific Self-Efficacy construct.

The active learning IMU intervention may also impact student intention to persist in engineering, and given that no well-established intention to persist instrument exists [442], the three items from the original LAESE instrument are used. Though

not explicitly identified as a construct, three items on the LAESE assessed student intention to persist (referred to as "retention items" in [308]). Four additional items were included in the modified version of the instrument to create a Student Affect construct.

The objective of this study is to confirm the validity and reliability of the four constructs before proceeding to evaluating the effects of the active learning IMU intervention on each of them. To that end, Fig. 10.1 summarizes the source of the items associated with each of the four constructs within the modified version of the LAESE instrument selected for this study. In addition, Table 10.1 reports each item on the modified LAESE. Thus, the instrument includes six items measuring Engineering Self-Efficacy, three items measuring Course-Specific Self-Efficacy, three items measuring Intention to Persist, and four items to assess Student Affect.

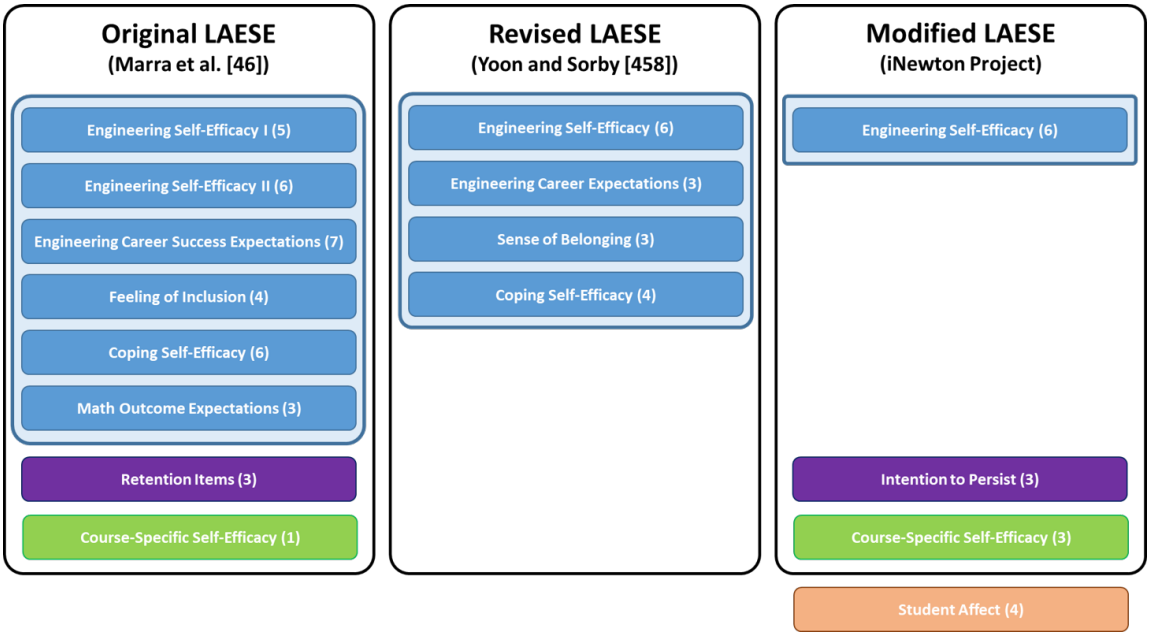


Figure 10.1: Original, revised, and modified versions of the LAESE. Boxes designate constructs (number of items). Blue boxes are those designed by the original creators. Purple and green boxes denote items included in the original version, but not as stand-alone constructs by themselves. The orange box represents an added new construct.

Table 10.1: Items on the modified LAESE for each of the four constructs: Engineering Self-Efficacy (ESE), Course-Specific Self-Efficacy (CSSE), Intention to Persist (PER), and Student Affect (SA).

Construct	No.	Item
ESE		
	1	I can succeed in an engineering curriculum.
	2	I can complete the math requirements for most engineering majors.
	3	I can complete the physics requirements for most engineering majors.
	4	I can do well in an engineering major during the current academic year.
	5	I can complete any engineering degree at this institution.
	6	Someone like me can succeed in an engineering career.
CSSE		
	7	I am confident that I can earn an A or B in this course.
	8	I am confident that I understand the concepts presented in this course.
	9	I am confident that I can apply the concepts from this course to a problem in the future.
PER		
	10	At the present time, how confident are you that you will keep your intended major through college?
	11	At the present time, how confident are you that you will be enrolled in any major in science or engineering in the next academic year?
	12	At the present time, how confident are you that you will graduate with your intended major?
SA		
	13	The experiments improved my learning experience.
	14	The experiments increased my interest in the subject matter.
	15	The experiments were engaging.
	16	The experiments increased my confidence in running my own experiments.

Methods

The data for the aforementioned modified version of the LAESE were collected at the end of the course for four semesters in this study (i.e., Fall 2017 onward). Refer to Chapter 8 for the semesters encompassed by this study and related demographic data. Furthermore, students who did not provide answers to all 16 items were eliminated from consideration. Consequently, 584 observations (nearly 20% more than the number of samples used by Yoon and Sorby [459] for their factor analyses) are included in the analysis described next.

Given the thorough investigation described and conducted by Yoon and Sorby [459], only a confirmatory factor analysis was conducted to substantiate the validity of the latent factor structure of the modified version of the LAESE used in this research. Specifically, there are 16 items mapping onto 4 constructs, namely: 1) Engineering Self-Efficacy, 2) Course-Specific Self-Efficacy, 3) Intention to Persist, and 4) Student Affect. The confirmatory factor analysis is conducted and evaluated using the R package "lavaan" [460]. Based on the fit indexes provided by that package, the residual mean square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), and standardized root mean square residual (SRMR) were used to judge the fit of the confirmatory factor analysis model [461]. The model fit indexes are considered in the acceptable range when RMSEA is close to 0.06 or less, CFI and TLI values are close to 0.95 or greater, and the SRMR is less than 0.08 [458, 461]. For reliability, internal consistency reliability coefficients were obtained for the items identified for the confirmed constructs. Acceptable values for Cronbach's alpha are typically greater than 0.7, though measures between 0.8 and 0.9 are desirable [397].

Results and Discussion

First, the correlation coefficients among the four constructs ranged from 0.11 to 0.77 as shown in Table 10.2, which suggests no significant multicollinearity exists between constructs [458]. The strongest correlation was between Engineering Self-Efficacy and Intention to Persist, and the weakest correlation was between Course-Specific Self-Efficacy and Student Affect.

Table 10.2: Correlation coefficients among the four constructs: Engineering Self-Efficacy (ESE), Course-Specific Self-Efficacy (CSSE), Intention to Persist (PER), and Student Affect (SA).

Construct	**ESE**	**CSSE**	**PER**	**SA**
ESE	1.00	0.24	0.77	0.24
CSSE		1.00	0.22	0.11
PER			1.00	0.26
SA				1.00

All fit indexes from the confirmatory factor analysis indicated the latent factor structure for the constructs described in Table 10.1 was acceptable: $\chi^2(98)=271.6$, $p<0.001$, RMSEA=0.055, CFI=0.959, TLI=0.950, and SRMR=0.046. The 90% confidence interval for RMSEA was [0.047, 0.063], which is well within the acceptable range of values. Table 10.3 reports the factor loadings for the model, all of which were statistically significant. Cronbach's alpha for Engineering Self-Efficacy, Course-Specific Self-Efficacy, Intention to Persist, and Student Affect were 0.76, 0.76, 0.77, and 0.92, respectively. The overall Cronbach's alpha for the modified LAESE is 0.83. These results indicate acceptable reliability in the students' responses for these constructs [462].

Conclusions

The first objective of this study was to validate that the four constructs (Engineering Self-Efficacy, Course-Specific Self-Efficacy, Intention to Persist, and Student Affect) contained within the modified version of the LAESE are distinguishably different from each another. The second objective was to confirm the internal reliability of the four constructs as well as the modified LAESE as a whole. The results from the confirmatory factor analysis, which includes the correlation coefficients among constructs and the item factor loadings contributing to their corresponding constructs, indicate that the constructs are distinct and their respective items are all contributing meaningfully to them. The Cronbach's alpha for the four constructs individually as well as the overall instrument indicate it is reliable as well. Thus, the four con-

structs are valid and reliable for investigating the effects of the active learning IMU intervention on self-efficacy (engineering and course-specific), intention to persist, and student affect as pursued in the remaining three studies.

Table 10.3: Unstandardized and standardized parameter estimates (FL: factor loadings and SE: standard errors) for the items contributing to each of the four constructs: Engineering Self-Efficacy (ESE), Course-Specific Self-Efficacy (CSSE), Intention to Persist (PER), and Student Affect (SA).

		Unstandardized		**Standardized**	
Construct	**No.**	**FL**	**SE**	**FL**	**SE**
ESE					
	1	1.00	NA	0.69	0.03
	2	0.87	0.07	0.58	0.03
	3	0.97	0.07	0.63	0.03
	4	1.14	0.07	0.77	0.02
	5	1.33	0.13	0.50	0.04
	6	0.79	0.07	0.55	0.03
CSSE					
	7	1.00	NA	0.85	0.03
	8	0.26	0.03	0.37	0.04
	9	1.09	0.07	0.94	0.03
PER					
	10	1.00	NA	0.75	0.03
	11	0.79	0.06	0.60	0.03
	12	0.95	0.06	0.82	0.02
SA					
	13	1.00	NA	0.88	0.01
	14	1.08	0.03	0.93	0.01
	15	0.95	0.03	0.85	0.01
	16	0.89	0.04	0.80	0.02

All factor loadings were statistically significant with $p<0.05$.

10.2.2 Study 2: Active Learning IMU Intervention and Self-Efficacy

Study Background and Objective

The second study leverages the findings from the first study to investigate the effects of the active learning IMU intervention on students' self-efficacy, both engineering and course-specific. Included in this investigation is a comparison with a Control group (Level 0) who did not engage with any implementations of the intervention, the levels of which are: 1) Demonstrations, 2) Prescribed Experiments, and 3) Student Projects; refer to Section 8.1. These three levels increase student engagement with the IMU technology, thereby providing increasingly more authentic *mastery experiences* that have the potential to build students' self-efficacy. However, it is currently unclear how the levels of the intervention will affect the two different types of self-efficacy given that Engineering Self-Efficacy relates to the students' perceptions of their abilities in a general engineering curriculum whereas Course-Specific Self-Efficacy pertains to their perceptions of their abilities relating to ME240.

Methods

As with the investigation into the intervention effects on conceptual understanding (Section 9.2.3), regression analyses are used to investigate the effects of the intervention rather than gains. As explained by Theobald and Freeman, gains alone lack the ability to distinguish if improvements are the result of an intervention or differences in student characteristics [407]. With regression analyses, differences in self-efficacy at the beginning of the semester can be accounted for by treating the Pre-score values as a predictor variable. Regression analyses also facilitate investigations into differences due to gender, ethnicity, or other factors as well as investigations into the interactions between them. Table 10.4 specifies each outcome and predictor variables as well as the different groups for the categorical variables. The Post outcome variable and Pre predictor variable refers to the Post- and Pre- scores for either Engineering Self-Efficacy (ESE) or Course-Specific Self-Efficacy (CSSE).

Referring to the student demographics described by Table 8.2, White students are those who identified themselves as Caucasian but not of Hispanic origin and Asian students are those who specifically identified themselves as Asian. Students

Table 10.4: Summary of multiple linear regression variables. Post and Pre (referring to Engineering Self-Efficacy or Course-Specific Self-Efficacy) are continuous variables whereas the rest are categorical variables.

Type	Variable	Categories
Outcome		
	Post	
Predictor		
	Pre	
	Intervention	Level 0, Level 1, Level 2, Level 3
	Gender	Male, Female
	Ethnicity	White, Asian, URM
	Section Type	Large, Small
	Major	ME, AERO, NAME, Elective

in the URM (underrepresented minorities) category are those who identified as Black/African American, Hispanic/Latino, American Indian/Alaskan Native, Native Hawaiian/Pacific Islander, or two or more. The Small sections enroll fewer than 20 students whereas Large sections enroll significantly more than 20 students. Students who were required to take ME240 as part of their major were in Mechanical Engineering (ME), Aerospace Engineering (AERO), or Naval Architecture and Marine Engineering (NAME). Students who took the class as a technical elective (i.e. did not declare one of those three majors) were broadly categorized as an "Elective" group.

Using the variables described in Table 10.4, the final multiple linear regression model used for Engineering Self-Efficacy (ESE) is

$$Post \sim 1+Intervention^{*}(Pre+Gender+Ethnicity)+SectionType+Major \quad (10.1)$$

where "1" is a potentially significant intercept and the asterisk indicates interactions between predictor variables are included in the model.

Using an Analysis of Variance (ANOVA), it was determined that the interactions

between the intervention and CSSE Pre-score, gender, and ethnicity did not significantly improve the model ($F(12, 948) = 0.57, p = 0.87$), so those interactions are removed from the model. However, in considering the joint distributions for Pre- and Post-scores by different student demographics, a potentially significant interaction between CSSE Pre-score and both gender and ethnicity was revealed. An ANOVA provides support for this assertion as well ($F(3, 957) = 2.60, p = 0.05$). Thus, the final linear regression model used for Course-Specific Self-Efficacy (CSSE) is

$$Post \sim 1 + Intervention + Pre^{*}(Gender + Ethnicity) + SectionType + Major \quad (10.2)$$

where "1" is a potentially significant intercept and the asterisk indicates interactions between predictor variables are included in the model.

It is assumed that each students' response is independent of other students' responses. For the results that follow, normality and heteroscedasticity assumptions are confirmed for the residuals for the models.

Results and Discussion

Table 10.5 contains the full results from the linear regression analysis conducted for ESE, represented in Eqn. 10.1. The model was significant ($F(23, 962) = 15.6$, $p < 0.001$), indicating that it is a significantly better fit than an intercept-only (average-only) model. Overall, the model accounts for 25% of the variation in the ESE Post-scores ($R^2 = 0.25$). Similarly, Table 10.6 contains the full results from the linear regression for CSSE. This model was also significant ($F(14, 957) = 20.0$, $p < 0.001$), indicating that it is a significantly better fit than an intercept-only (average-only) model. Overall, it describes 21% of the variation in the CSSE Post-scores ($R^2 = 0.21$). In both tables, CI:LB and CI:UB denote the lower and upper bounds of the confidence intervals for the estimated coefficient. The variables in the parentheses are the baseline (comparison) categories. For interpretation of the results in the tables, all items were on a 7-point Likert scale that were normalized to 1. A value of 0 corresponds to low self-efficacy, 0.5 corresponds to neutral self-efficacy, and 1 corresponds to high self-efficacy.

Table 10.5: Summary of ESE multiple linear regression results.

Variable	Coefficient (β)	CI:LB	CI:UB	t-value	p-value
(Intercept)	0.56	0.42	0.70	8.0	$< 0.001^{\ddagger}$
Pre	0.35	0.20	0.50	4.5	$< 0.001^{\ddagger}$
Intervention (Level 0)					
Level 1	-0.22	-0.39	-0.04	-2.4	0.01^{*}
Level 2	-0.21	-0.38	-0.04	-2.4	0.02^{*}
Level 3	-0.32	-0.50	-0.13	-3.4	$< 0.001^{\ddagger}$
Gender (Male)					
Female	-4e-3	-0.05	0.04	-0.2	0.88
Ethnicity (White)					
Asian	-7e-3	-0.04	0.03	-0.4	0.68
URM	-3e-3	-0.06	0.05	-0.1	0.92
Major (ME)					
AERO	0.01	-0.01	0.02	1.3	0.20
NAME	-0.01	-0.04	0.02	-0.5	0.65
Elective	0.01	-0.02	0.03	0.4	0.65
Section Type (Large)					
Small	1e-3	-0.03	0.03	0.1	0.93
Interactions					
Level 1*Pre	0.25	0.05	0.44	2.5	0.01^{*}
Level 2*Pre	0.22	0.04	0.41	2.3	0.02^{*}
Level 3*Pre	0.35	0.15	0.56	3.4	$< 0.001^{\ddagger}$
Level 1*Female	-0.02	-0.07	0.04	-0.6	0.57
Level 2*Female	3e-4	-0.05	0.05	0.01	0.99
Level 3*Female	3e-3	-0.05	0.06	0.1	0.92
Level 1*Asian	-0.01	-0.05	0.04	-0.3	0.79
Level 2*Asian	-0.02	-0.06	0.03	-0.7	0.51
Level 3*Asian	-7e-4	-0.05	0.05	0.03	0.98
Level 1*URM	0.01	-0.06	0.07	0.2	0.86
Level 2*URM	0.01	-0.06	0.07	0.2	0.83
Level 3*URM	-0.02	-0.09	0.05	-0.6	0.58

Significant $\alpha = {}^{\bullet}0.1$, ${}^{*}0.05$, ${}^{\dagger}0.01$, and ${}^{\ddagger}0.001$.

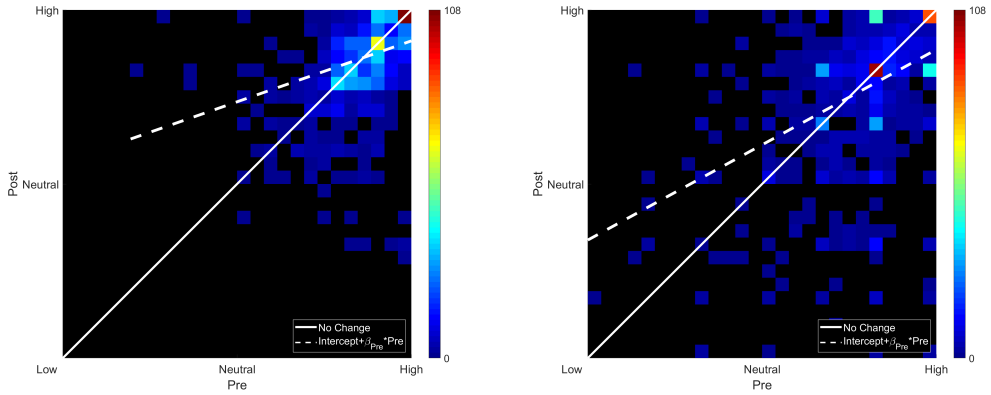
Table 10.6: Summary of CSSE multiple linear regression results.

Variable	Coefficient (β)	CI:LB	CI:UB	t-value	p-value
(Intercept)	0.34	0.25	0.43	7.5	$< 0.001^{\ddagger}$
Pre	0.55	0.44	0.65	10.6	$< 0.001^{\ddagger}$
Intervention (Level 0)					
Level 1	0.02	-0.01	0.06	1.2	0.24
Level 2	0.01	-0.03	0.04	0.3	0.78
Level 3	0.05	0.01	0.09	2.4	0.02^{*}
Gender (Male)					
Female	0.09	-0.05	0.23	1.3	0.20
Ethnicity (White)					
Asian	-0.17	-0.31	-0.03	-2.4	0.02^{*}
URM	-0.13	-0.31	0.05	-1.4	0.17
Major (ME)					
AERO	-0.02	-0.05	0.01	-1.4	0.16
NAME	-0.07	-0.12	-0.01	-2.4	0.02^{*}
Elective	-0.03	-0.08	0.02	-1.2	0.22
Section Type (Large)					
Small	-0.3	-0.09	0.02	-1.3	0.19
Interactions					
Pre*Female	-0.16	-0.34	0.01	-1.8	$0.07^{\bullet}$
Pre*Asian	0.18	0.01	0.35	2.1	0.04^{*}
Pre*URM	0.14	-0.08	0.36	1.2	0.22

Significant $\alpha = {}^{\bullet}0.1$, ${}^{*}0.05$, ${}^{\dagger}0.01$, and ${}^{\ddagger}0.001$.

The significant intercepts and Pre-score coefficients describe on average how self-efficacy (engineering or course-specific) changed over a semester regardless of intervention level or specific student demographics. For example, the ESE intercept means a student who had very low (0) ESE at the beginning of the semester had at least neutral (0.56) self-efficacy at the end of the semester. The significant ESE Pre-score coefficient means students gained on average 35% of their Pre-score ESE

by the end of the semester. The CSSE intercept means a student who had very low (0) CSSE at the beginning of the semester had slightly low (0.34) self-efficacy at the end of the semester. The significant CSSE Pre-score coefficient means students gained on average 55% of their Pre-score ESE by the end of the semester. Figure 10.2 illustrates how these two predictor variables relate to the baseline case (no change) for each type of self-efficacy. The majority of both the ESE and CSSE joint distributions is concentrated in the upper right corner of the plots, indicating a significant proportion of students entered and exited ME240 with high self-efficacy.



(a) (b)

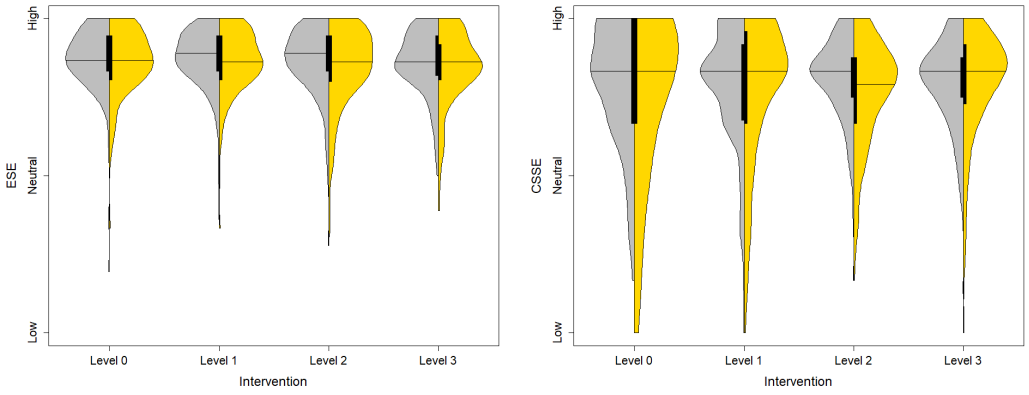
Figure 10.2: The background is a heat map of the joint distribution of Pre- and Post-scores for **(a)** ESE and **(b)** CSSE (black areas are void of data). The solid white line denotes the baseline case when the Pre- and Post-scores are equal (no change in self-efficacy as measured by the LAESE). The dashed line illustrates significant intercepts and the significant slopes relating Pre- and Post-scores.

Students appear to have higher ESE than CSSE, which could be the result of the fact that they are evaluated on a weekly basis on their performance in ME240. The end of the semester survey is administered after the last day of class, but before the cumulative final. As such, students likely have a realistic sense for the strength of their abilities in ME240. However, it is unclear what (if any) kind of feedback students receive with respect to their abilities as engineers in their major's curriculum. It could be that other classes taken concurrently with ME240 could be providing op-

portunities to authentically practice their engineering skills (i.e. laboratory classes). Discussions of the results for the remaining predictor variables in Tables 10.5 and 10.6 follow. While discussions pertain to a subset of predictor variables, the results presented are with respect to controlling for variations due to the other variables.

Intervention Level

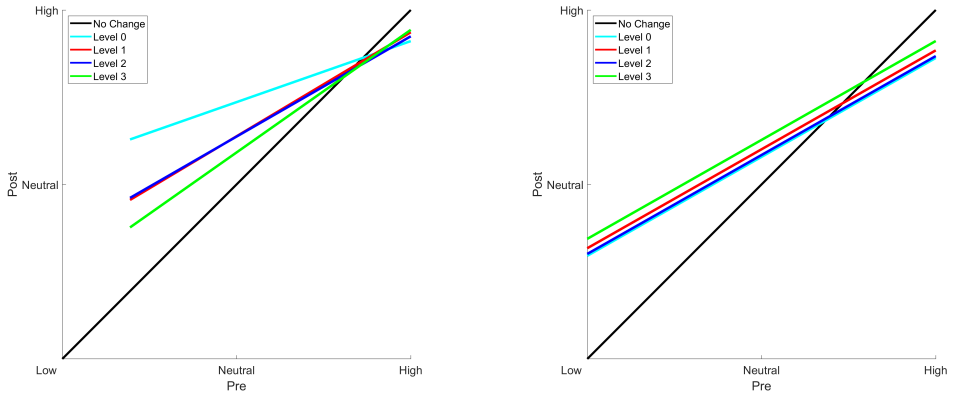
Figure 10.3 contains split violin plots describing each intervention level's Pre- and Post-score distributions for ESE and CSSE. Each plot presents two separate distributions: the distributions for the Pre-scores are on the left (grey), and the one for the Post-scores is on the right (gold). The black boxes at the middle of each side of the violin are box plots corresponding to each distribution with the black horizontal lines indicating the median of that distribution. With respect to the medians in Fig. 10.3a, Levels 1 (Demonstrations) and 2 (Prescribed Experiments) may be associated with a decrease in ESE as compared to Level 0 (Control) and Level 3 (Student Projects), which appear not to change ESE. With respect to the medians in Fig. 10.3b, only Level 2 appears to be associated with decreases in CSSE as compared to the other levels, which do not appear to have any effect.



(a) (b)

Figure 10.3: Split violin plots by intervention level of the Pre- and Post-score distributions (grey and gold, respectively) for **(a)** ESE and **(b)** CSSE. The distributions are for each of the intervention levels: Level 0 (Control), Level 1 (Demonstrations), Level 2 (Prescribed Experiments), and Level 3 (Student Projects).

As evidenced by the differences in the distributions in Fig. 10.3, there are qualitative differences in ESE and CSSE by intervention level. Figure 10.4 illustrates these differences by intervention level. This includes the significant interactions between intervention and ESE Pre-score whereas the final CSSE regression model (Eq. 10.2) did not include these interactions.



(a) (b)

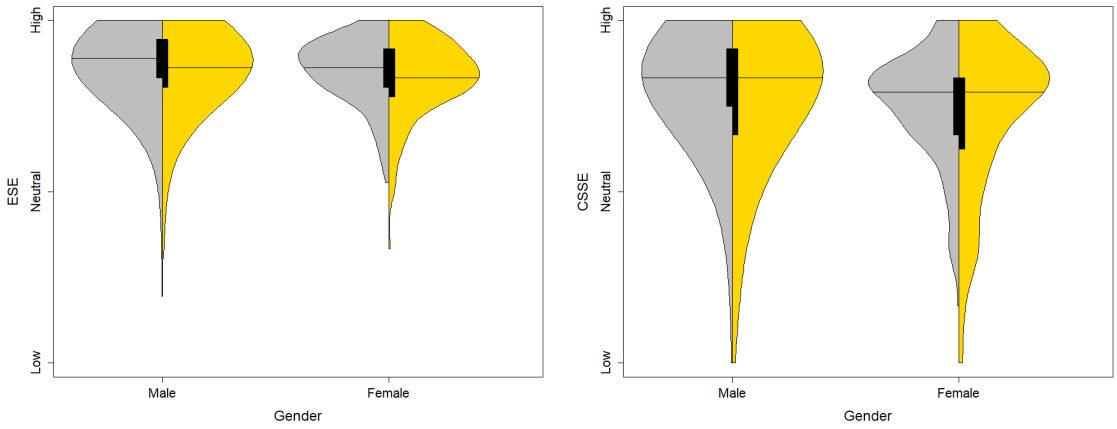
Figure 10.4: Results of the regression analyses by intervention level for **(a)** ESE and **(b)** CSSE. The black line denotes the baseline case that there was no change over the course of the semester. The four other lines distinguish each of the intervention levels: Level 0 (Control), Level 1 (Demonstrations), Level 2 (Prescribed Experiments), and Level 3 (Student Projects).

From the regression results, ESE improves less for students who are more engaged with the active learning IMU intervention. However, the significant interactions between intervention and Pre-scores indicates that the different levels of the intervention disproportionately affect students based on their ESE at the beginning of the semester. For example, the Student Projects group (Level 3) has the most significant drop on average (-0.32) as well as the greatest interaction with Pre-score (0.35). For comparable start of semester ESE, more engagement with the technology is associated with lower ESE at the end of the semester, though this gap shrinks as the start of semester ESE is higher. This could be evidence that students who have developed a resilient sense of ESE are less deterred by the increase in cognitive

engagement required for the higher levels of the intervention. The opposite trend is observed for CSSE; on average, CSSE improved for students who participated in the Student Projects. This is expected as this level of the intervention provides students with opportunities to apply what they are learning in ME240 to a relatively unstructured dynamic system of their choosing. During this process, they are also receiving feedback on how to interpret their results and improve their experiments, which likely further solidifies their CSSE. The difference in results between ESE and CSSE could indicate that there are other factors contributing to ESE (and possibly interacting with their experiences engaging with the intervention) that are not included in the current analyses.

Gender

Next, the distributions for Pre- and Post-scores for ESE and CSSE with respect to gender are illustrated in Fig. 10.5 with split violin plots. Noting the distributions' medians, both male and female students appear to experience a drop in ESE of approximately the same magnitude whereas differences in CSSE from the beginning to the end of a semester in both genders appear nominal. The results from the multiple regression are illustrated in Fig. 10.6.



(a) (b)

Figure 10.5: Split violin plots by gender of the Pre- and Post-score distributions (grey and gold, respectively) for **(a)** ESE and **(b)** CSSE.

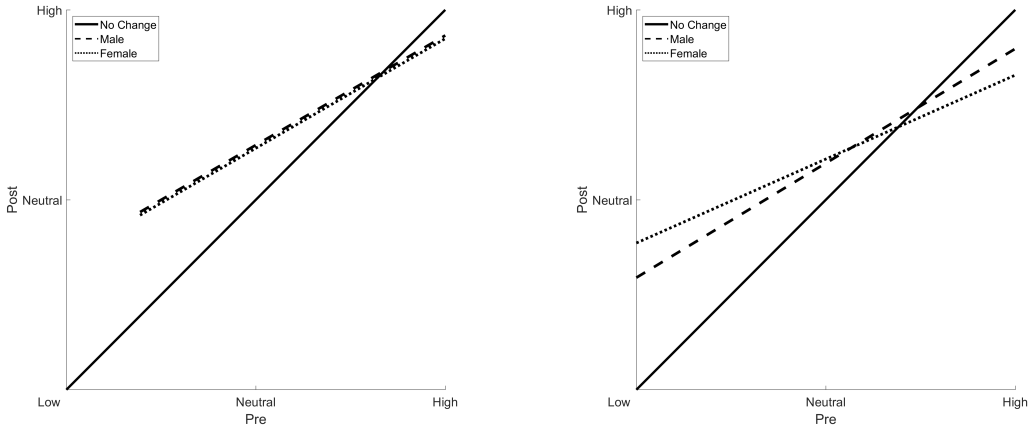


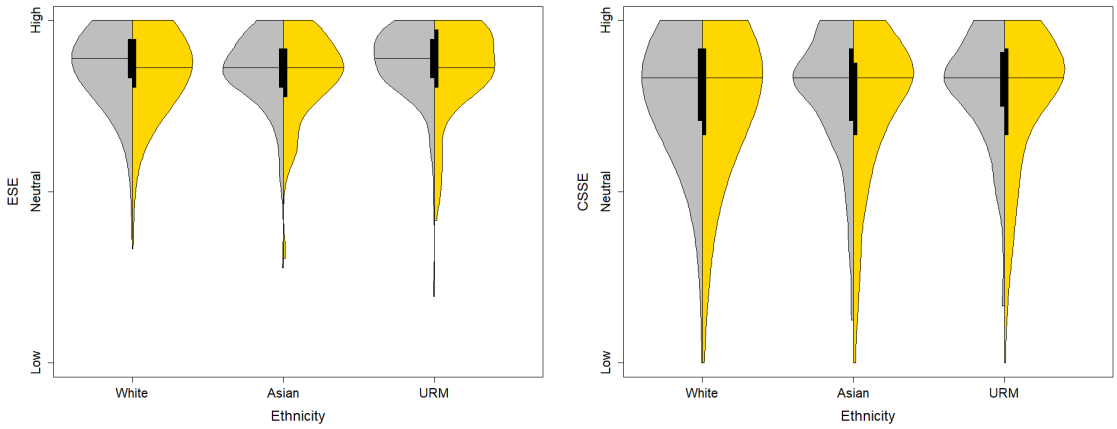
Figure 10.6: Results of the regression analyses by gender for **(a)** ESE or **(b)** CSSE. The black line denotes the baseline case that there was no change over the course of the semester. The other two lines distinguish gender.

As evidenced by the largely overlapping distributions in Fig. 10.5, ESE and CSSE differences by gender are not significant. The final ESE regression model (Eq. 10.1) included interactions between intervention and gender, which were also not significant. However, the final CSSE regression model (Eq. 10.2) included a noteworthy interaction between gender and Pre-score ($\beta = -0.16$, $p = 0.07$). Female students with higher initial CSSE had larger decreases at the end of the semester as compared to their male peers. This result is partially explained by the results reported by Brown and Matusovich [309], who found female students report lower self-efficacy despite comparable academic performance. In Chapter 9, the differences in conceptual understanding as measured by the DCI between male and female students were found to be statistically nonsignificant. The results of that chapter also found performance on the DCI is representative of performance on a midterm examination. Using DCI performance as a proxy for academic performance in ME240, female students were reporting lower CSSE at the end of the semester despite comparable academic performance. It is unclear if female students were underestimating their abilities, male students were overestimating their abilities, or potentially both

groups of students have a mismatch between their self-efficacy and abilities.

Ethnicity

The Pre- and Post-scores distributions for ESE and CSSE by ethnicity are illustrated in 10.7. With respect to the distributions' medians, White and URM students appear to have decreases in ESE over the semester whereas Asian students seem to stay relatively constant. The CSSE distributions' medians do not indicate there are any differences for any group.



(a) (b)

Figure 10.7: Split violin plots by ethnicity of the Pre- and Post-score distributions (grey and gold, respectively) for **(a)** ESE and **(b)** CSSE.

The results from the multiple regression analyses are illustrated in Fig. 10.8. The differences in ESE for each group were not significant nor were the interactions between intervention and ethnicity. However, there were significant differences in CSSE by ethnicity, particularly for Asian students ($\beta = -0.17$, $p = 0.02$). The CSSE regression model (Eq. 10.2) also included an interaction between Asian students and their Pre-scores ($\beta = 0.18$, $p = 0.04$). When considering the results in Fig. 10.8, Asian students with lower levels of CSSE at the beginning of the semester report significantly lower levels at the end of the semester as compared to White students. There is some support for this finding in the literature. For example,

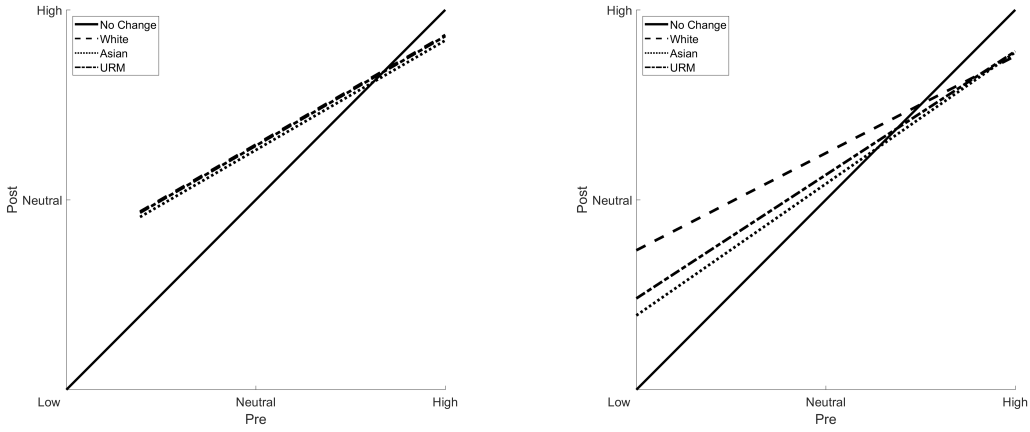


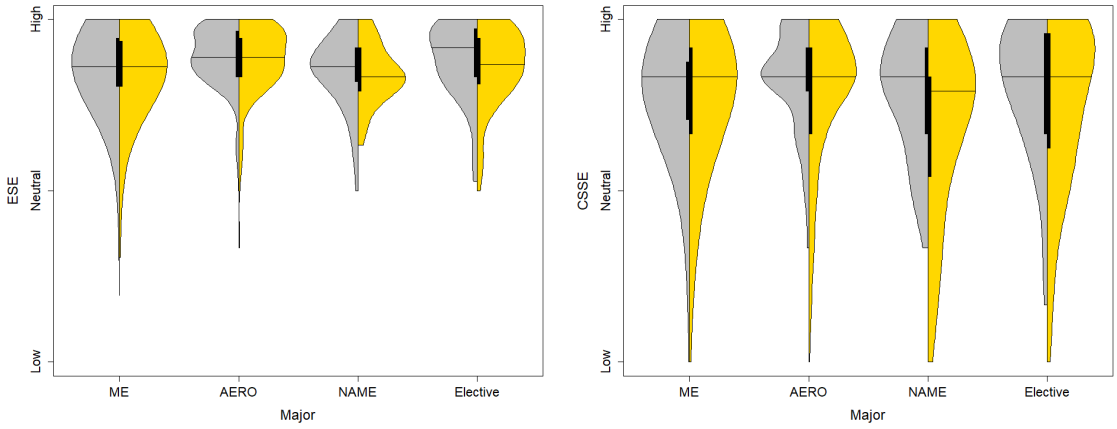
Figure 10.8: Results of the regression analyses by ethnicity for **(a)** ESE and **(b)** CSSE. The black line denotes the baseline case that there was no change over the course of the semester. The other three lines distinguish ethnicity.

Eaton and Dembo [463] found Asian American 9th grade students reported lower self-efficacy than their non-Asian counterparts. In Fig. 10.8, it appears that URM students, particularly those with lower CSSE at the beginning of the semester, may also have a more significant drop by the end of the semester as compared to White students. From Table 10.6, the coefficient for URM students (-0.13) has a comparable magnitude as the coefficient for Asian students (-0.17), but the difference is not statistically significant ($p = 0.17$). This result is partially explained by the large variations in Fig. 10.7 and the unequal sample sizes of URM ($N = 130$) students compared to White ($N = 657$) and Asian ($N = 203$) students in addition to the large variation in the Post-score distribution for URM students in Fig. 10.8.

Major

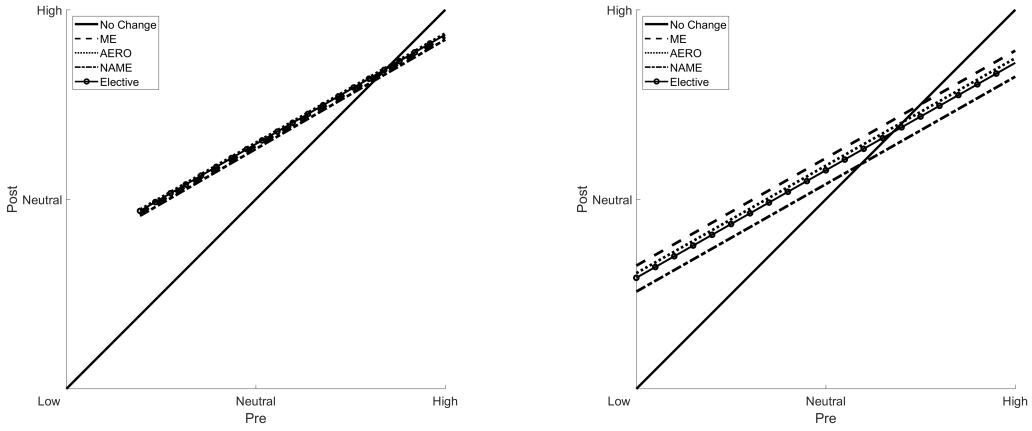
The Pre- and Post-score distributions for ESE and CSSE by major are illustrated in Fig. 10.9. Observation of the distributions' medians reveals students taking ME240 as an elective or students in the NAME department may experience decreases in ESE over the course of a given semester. Students in the NAME department also appear to have more significant decreases from the beginning to the end of a semester

in CSSE as compared to students in other majors. The results from the regression analysis are illustrated in Fig. 10.8.



(a) (b)

Figure 10.9: Split violin plots by major of the Pre- and Post-score distributions (grey and gold, respectively) for **(a)** ESE and **(b)** CSSE.



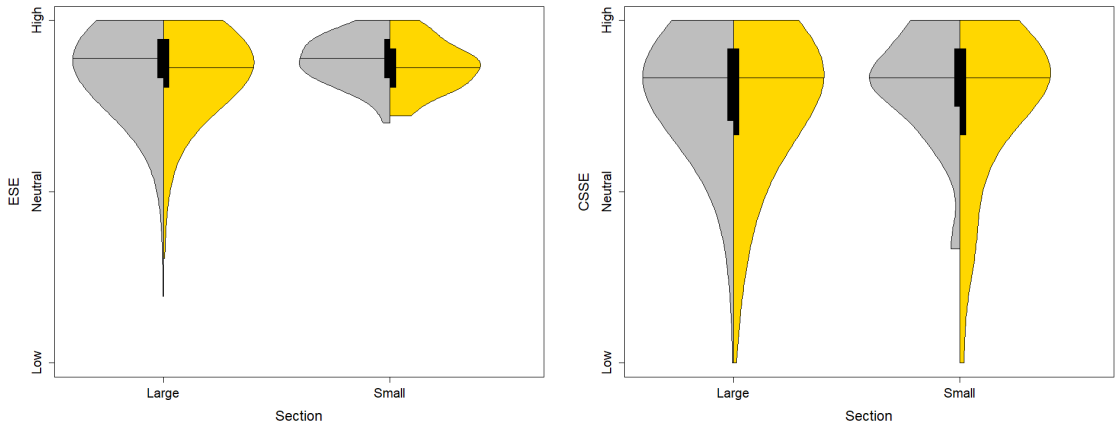
(a) (b)

Figure 10.10: Results of the regression analyses by major for **(a)** ESE and **(b)** CSSE. The black line denotes the baseline case that there was no change over the course of the semester. The other four lines distinguish major.

The regression analyses revealed no significant differences in ESE by major. However, there is a significant difference in CSSE for NAME students, who on average report lower levels of CSSE at the end of the semester as compared to ME students. Given that the results of Chapter 9 detected no significant differences in conceptual understanding as measured by the DCI between NAME and ME students, the reason for this difference can only be answered with a follow-up qualitative study.

Section Type

The different section sizes are the result of a small section initiative at the University of Michigan, which began with the Provost's Office in the form of a request for proposal from major units. The College of Engineering responded by identifying via department input potential "gateway" courses where students who frequently appear to struggle may benefit from offerings of small sections (i.e., enrollments fewer than 20 students). The ME department identified their sophomore engineering science courses (including ME240) as those gateway courses, and began offering small sections for these courses intermittently starting in the Fall 2016 semester (i.e., the same semester this current study began). The distributions for Pre- and Post-scores for ESE and CSSE with respect to section type are illustrated in Fig. 10.11.



(a) (b)

Figure 10.11: Split violin plots by section type of the Pre- and Post-score distributions (grey and gold, respectively) for **(a)** ESE and **(b)** CSSE.

The regression results are illustrated in Fig. 10.12. No significant differences in ESE or CSSE were detected.

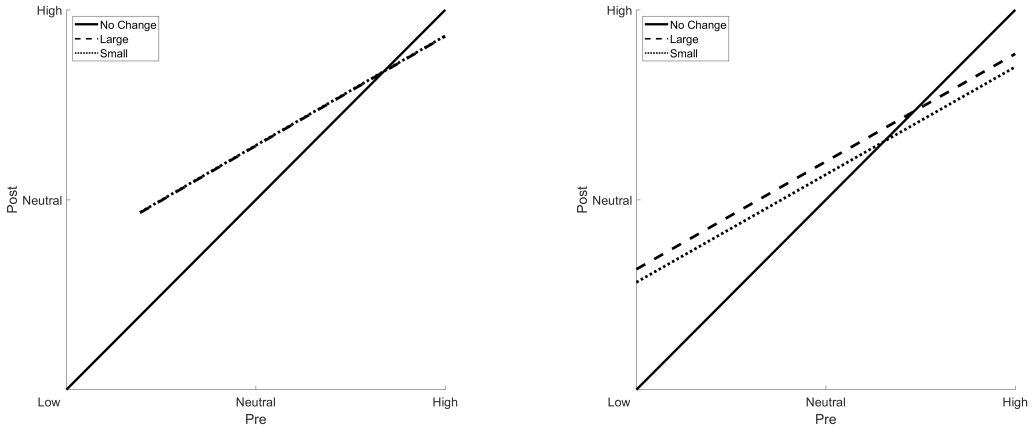


Figure 10.12: Results of the regression analyses by section type for **(a)** ESE and **(b)** CSSE. The black line denotes the baseline case that there was no change over the course of the semester. The other two lines distinguish section type.

Conclusions

The results from this second study into the impacts of the active learning IMU intervention on self-efficacy are mixed. Significant coefficients indicated ESE decreased with increasing levels of the intervention (increasing levels of engagement with the iNewton IMU technology). The significant interactions indicated that students who entered with higher levels of ESE at the start of the semester were less affected. However, students who participated in the Student Projects had a greater increase in CSSE by the end of the semester as compared to the students in all other intervention levels. It is interesting that the active learning IMU intervention appears to have opposing effects on the two types of self-efficacy included in this study. Given how many other factors could be contributing to ESE, it is possible that something else (i.e., experiences in another course) is interacting with their experiences during the active learning IMU intervention in ME240. Future work should consider this interaction as well as how students perceive different types of self-efficacy.

10.2.3 Study 3: Active Learning IMU Intervention and Intention to Persist

Study Background and Objective

The third study leverages the first study's findings to investigate the effects of the active learning IMU intervention on students' intentions to persist. Included in this investigation is a comparison with a Control group (Level 0) who did not engage with any implementations of the intervention. The introduction of active learning is expected to positively impact students' intention to persist by providing them with an engaging, authentic activity to practice their ME240 skills.

Methods

As with prior studies, regression analyses are used to investigate the effects of the active learning IMU intervention. Using the same variables in Table 10.4, the final regression model used for Intention to Persist (PER) is

$$Post \sim 1+Intervention^{*}(Pre+Gender+Ethnicity)+SectionType+Major \quad (10.3)$$

where "1" is a potentially significant intercept and the asterisk indicates interactions between predictor variables are included in the model. It is assumed that each students' response is independent of other students' responses. Normality and heteroscedasticity assumptions are confirmed for the residuals for the models.

Results and Discussion

Table 10.7 contains the full results from the regression analysis conducted for Intention to Persist (PER). The model was significant ($F(23, 966) = 14.8, p < 0.001$), indicating that it is a significantly better fit than an intercept-only (average-only) model. The model accounts for 24% of the variation in PER Post-scores ($R^2 = 0.24$). In the table, CI:LB and CI:UB denote the lower and upper bounds of the confidence intervals for the estimated coefficients. The variables in the parentheses are the baseline (comparison) categories. For interpretation, all PER items were on a 5-point Likert scale that were normalized to 1. A value of 0 corresponds to low PER, 0.5 corresponds to neutral PER, and 1 corresponds to high PER.

Table 10.7: Summary of PER multiple linear regression results.

Variable	Coefficient (β)	CI:LB	CI:UB	t-value	p-value
(Intercept)	0.35	0.20	0.49	4.7	$< 0.001^{\ddagger}$
Pre	0.63	0.47	0.78	7.9	$< 0.001^{\ddagger}$
Intervention (Level 0)					
Level 1	0.14	-0.03	0.30	1.6	0.12
Level 2	0.18	0.01	0.36	2.0	0.04^{*}
Level 3	0.22	0.05	0.39	2.5	0.01^{*}
Gender (Male)					
Female	-4e-3	-0.05	0.04	-0.2	0.85
Ethnicity (White)					
Asian	-0.03	-0.07	3e-3	-1.8	$0.07^{\bullet}$
URM	1e-3	-0.05	0.05	0.1	0.97
Major (ME)					
AERO	5e-3	-0.01	0.02	0.6	0.53
NAME	0.01	-0.02	0.04	0.9	0.36
Elective	-3e-3	-0.03	0.02	-0.2	0.83
Section Type (Large)					
Small	4e-3	-0.02	0.03	0.3	0.76
Interactions					
Level 1*Pre	-0.12	-0.31	0.06	-1.3	0.19
Level 2*Pre	-0.19	-0.38	5e-4	-2.0	$0.05^{\bullet}$
Level 3*Pre	-0.22	-0.41	-0.03	-2.3	0.02^{*}
Level 1*Female	0.01	-0.04	0.06	0.4	0.71
Level 2*Female	0.02	-0.03	0.08	0.9	0.38
Level 3*Female	0.02	-0.04	0.07	0.6	0.52
Level 1*Asian	0.01	-0.03	0.05	0.5	0.63
Level 2*Asian	0.03	-0.02	0.07	1.1	0.28
Level 3*Asian	1e-3	-0.05	0.05	0.1	0.96
Level 1*URM	-0.02	-0.08	0.05	-0.5	0.63
Level 2*URM	-0.01	-0.07	0.05	-0.3	0.73
Level 3*URM	-0.04	-0.11	0.02	-1.4	0.17

Significant $\alpha = {}^{\bullet}0.1$, ${}^{*}0.05$, ${}^{\dagger}0.01$, and ${}^{\ddagger}0.001$.

The significant intercept and Pre-score coefficient describe on average how students' PER changed over a semester regardless of intervention level or specific student demographics. The intercept means a student who had very low (0) PER at the beginning of the semester had at least moderately low (0.35) PER at the end of the semester. The significant Pre-score coefficient means students gained on average 63% of their Pre-score PER by the end of the semester. Figure 10.13 illustrates how these two predictor variables relative to the baseline case in which there is no change in PER over the course of the semester. The majority of joint distribution is concentrated in the upper right corner of the plots, indicating a significant proportion (nearly half) of students in this study entered and exited ME240 with high PER.

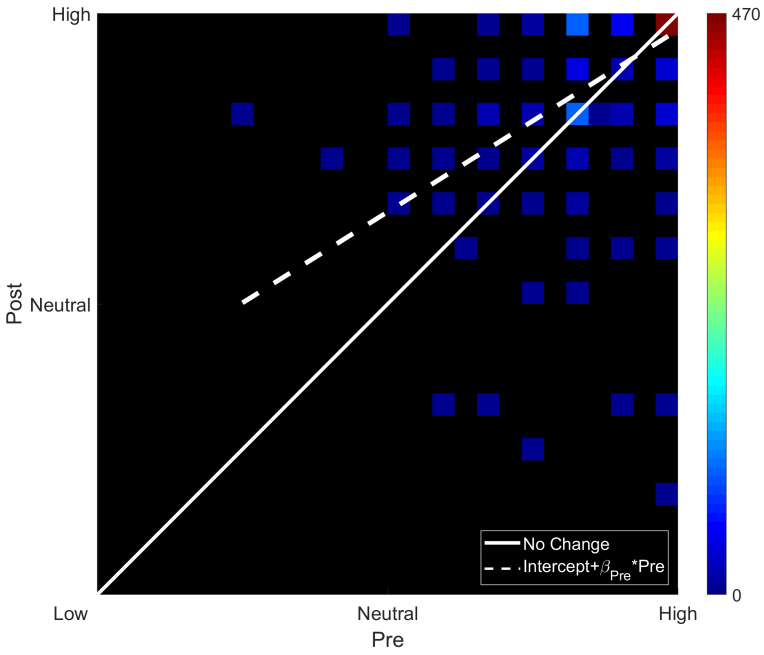


Figure 10.13: The background is a heat map of the joint distribution of Pre- and Post-scores (black areas are void of data). The solid white line denotes the baseline case when the Pre- and Post-scores are equal (no change in Intention to Persist). The dashed line illustrates the significant intercept (0.35) and the dotted line illustrates the significant intercept and the significant slope ($\beta_{Pre} = 0.63$) relating Pre- and Post-scores.

On average, it appears students' PER remained very high (1) or improved over the course of the semester. Discussions of the results for the remaining predictor variables in Table 10.7 follow next. While each discussion pertains to a subset of predictor variables, the results presented are with respect to controlling for variations due to the other variables.

Intervention Level

Figure 10.14 contains split violin plots illustrating the Pre- (grey) and Post-score (gold) distributions for PER for each intervention level. The black boxes at the middle of each violin are box plots for each distribution with the black horizontal lines indicating the median, which for most of the distributions is at the top of the violin. With the exception of the Pre-score distribution for Level 0, the medians are the top edge of the violins indicating that more than half of the students reported high PER. All of the distributions also appear to be bimodal with the majority

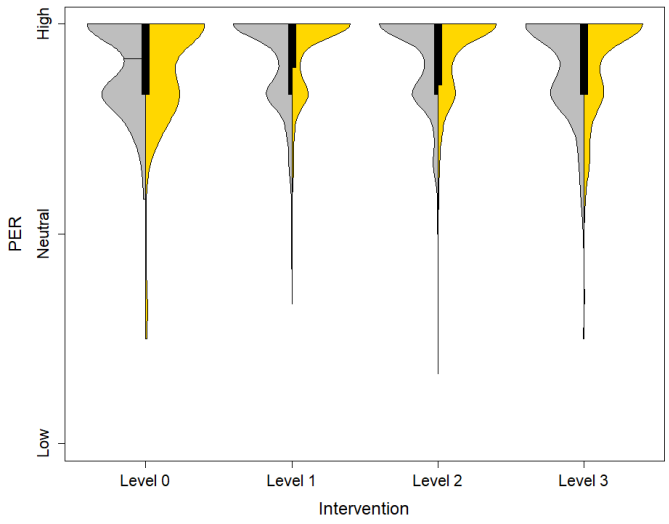


Figure 10.14: Split violin plots of the Pre- and Post-score distributions (grey and gold, respectively) for each of the intervention levels: Level 0 (Control), Level 1 (Demonstrations), Level 2 (Prescribed Experiments), and Level 3 (Student Projects).

of students reporting either very high or moderately high PER, likely due to the

fact that the PER items were on a 5-point Likert scale. However, the Post-score distributions (gold) for the Level 0 (Control) is noticeably different as compared to the other three levels, indicating there could be a significant difference in PER by intervention.

As evidenced by the qualitatively different distributions in Fig. 10.14, there are significant differences in PER by intervention level, particularly for Level 2 (Prescribed Experiments) and Level 3 (Student Projects). Figure 10.15 illustrates these differences by intervention level via the multiple regression results.

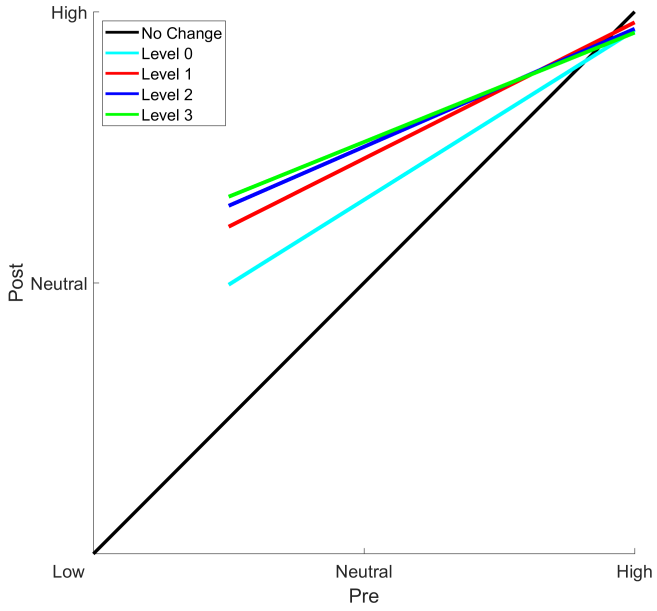


Figure 10.15: The black line denotes the baseline case when the Pre- and Post-scores are equal (no change in PER). The other lines represent the multiple regression results for each intervention level while controlling for variations in the other predictor variables: Level 0 (Control), Level 1 (Demonstrations), Level 2 (Prescribed Experiments), and Level 3 (Student Projects).

From Fig. 10.15, it appears that students entering ME240 with low levels of PER saw more significant gains as their engagement with the IMU technology increased. This result is meaningful and supported by the literature. For example, Graham

et al. [306] identified early research experiences and active learning in introductory classrooms in their Persistence Framework as two successful approaches in inspiring STEM students to persist in their respective majors. Given that the Student Projects implementation of the active learning IMU intervention is akin to a research experience, it is reasonable that the largest increases for students entering ME240 with lower levels of PER arise in Level 3. Additionally, Dweck [464] specifically noted the importance of high self-efficacy for persistence. The previous study revealed the Student Projects implementation of the active learning IMU intervention yielded a significant increase in Course-Specific Self-Efficacy. Thus, it makes sense that this implementation would also yield the greatest gains in PER.

Gender

The Pre- and Post-score distributions by gender in Fig. 10.16 are qualitatively different with the female students' distribution presenting as bimodal.

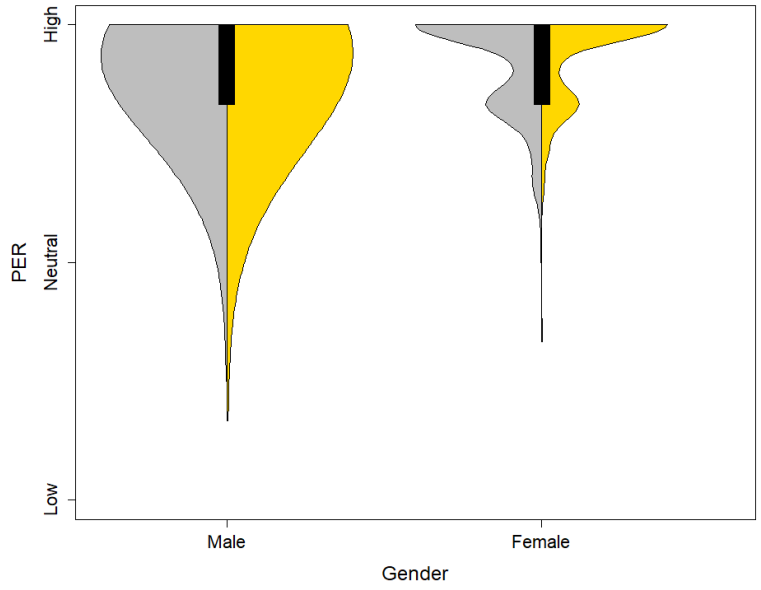


Figure 10.16: Split violin plots by gender of the Pre- and Post-score distributions (grey and gold, respectively) for PER.

Figure 10.17 illustrates the multiple regression results for PER by gender. Despite the qualitative differences in the distributions, the overall differences are not significant (consistent with the findings reported by Navarro et al. [465]), nor are the interactions between intervention level and gender. The appearance of the distributions in Fig. 10.16 is likely the result of the differing sample sizes with the male students (N=797) outnumbering the female students (N=227) approximately 4:1.

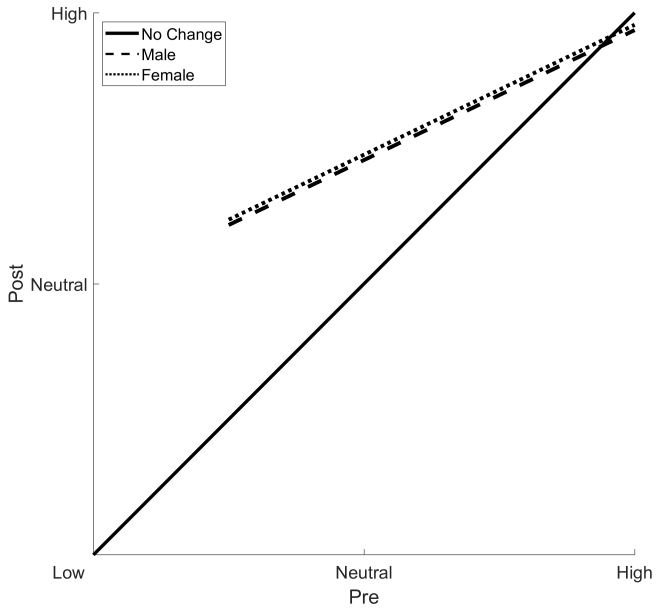


Figure 10.17: Results of the regression analyses by gender for PER. The black line denotes the baseline case that there was no change over the course of the semester. The other two lines distinguish gender.

Ethnicity

Figure 10.18 illustrates the Pre- and Post-score distributions by ethnicity. The differences the distributions' shape can be partially explained by the sample sizes of White students ($N = 657$) as compared to Asian ($N = 203$) and URM ($N = 130$) students. The Asian students' distributions appear more bimodal indicating a slightly lower PER on average. The regression results are illustrated in Fig. 10.19.

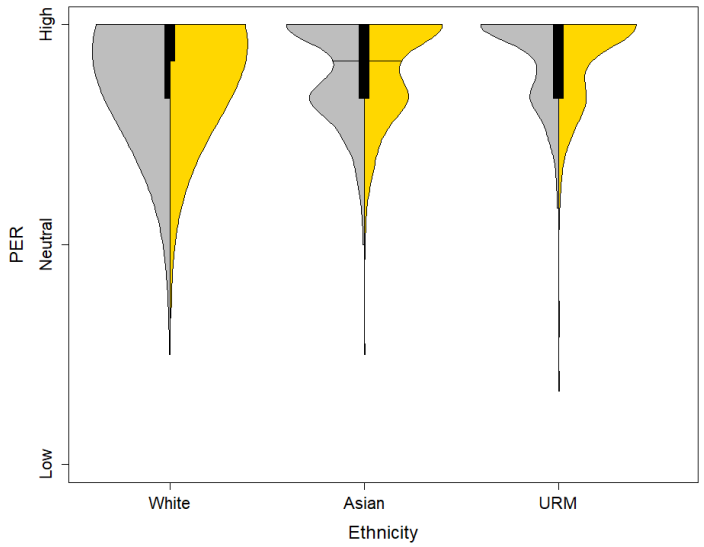


Figure 10.18: Split violin plots by ethnicity of the Pre- and Post-score distributions (grey and gold, respectively) for PER.

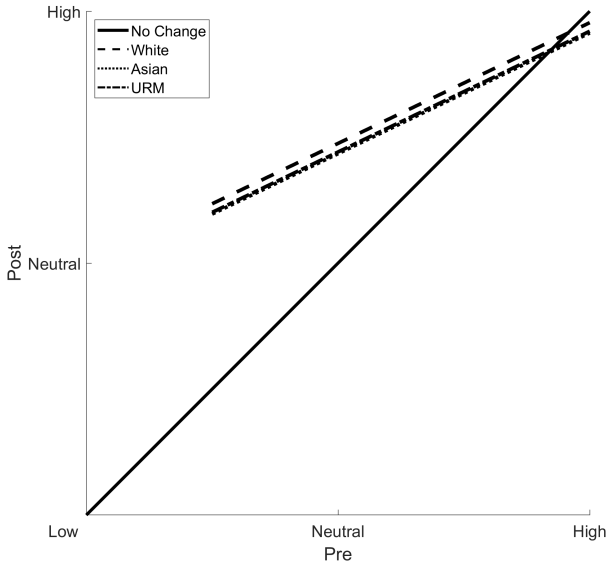


Figure 10.19: Results of the regression analyses by ethnicity for PER. The black line denotes the baseline case that there was no change over the course of a semester. The other three lines distinguish ethnicity.

Though the shape of the distributions appears qualitatively different, the regression results indicate these differences are largely not statistically significant, which is also consistent with the findings reported by Navarro et al. [465]. However, Asian students report marginally lower PER ($\beta = -0.03$, $p = 0.07$) as compared to White students. This decrease in PER could be related to the fact that Asian students consistently reported having lower Course-Specific Self-Efficacy (CSSE) at the end of the semester as compared to White students.

Major

The differences present in the Pre- and Post-score distributions for PER by major in Fig. 10.20 are not statistically significant according to the regression analysis, the results of which are illustrated in Fig. 10.21. Qualitatively, it appears students taking ME240 as a technical elective generally report lower PER and experience the largest (and positive) difference in PER over the course of a semester.

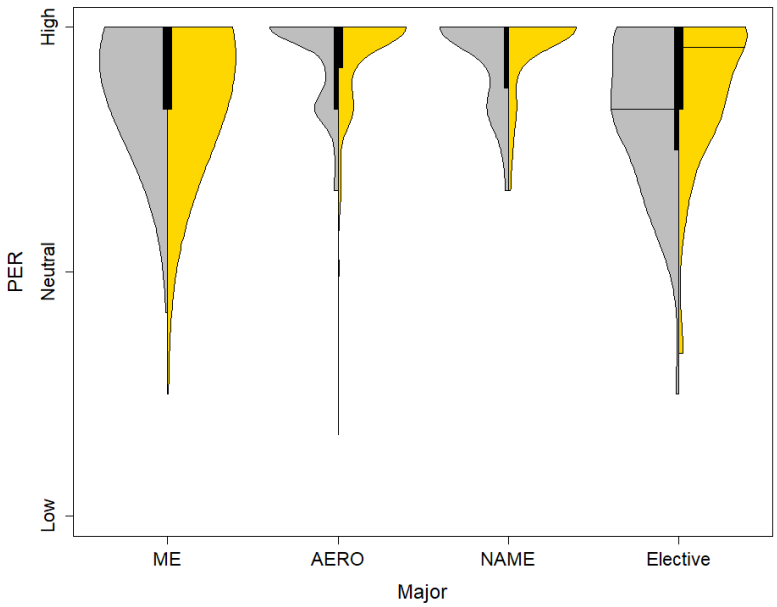


Figure 10.20: Split violin plots by major of the Pre- and Post-score distributions (grey and gold, respectively) for PER.

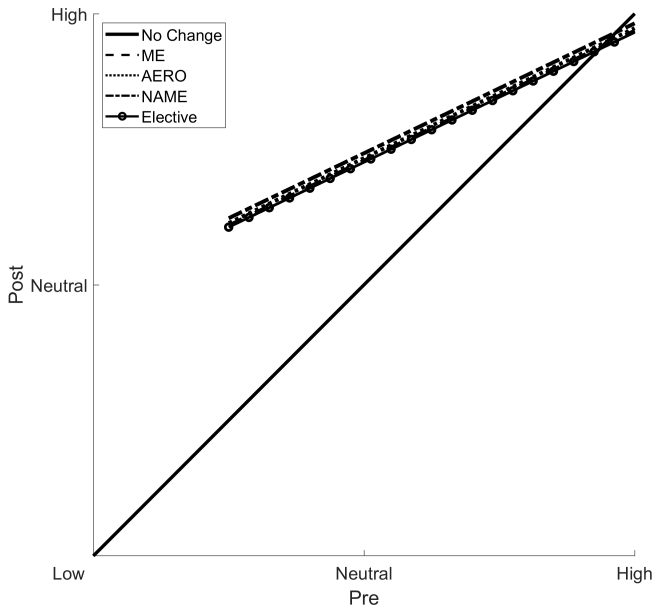


Figure 10.21: Results of the regression analyses by major for PER. The black line denotes the baseline case that there was no change over the course of the semester. The other four lines distinguish major.

The differences in the distributions in Fig. 10.20 are likely not significant as a result of the large variation and unequal sample sizes. For example, 669 students reported majoring in ME whereas only 58 students fell into the Elective group. Unlike Asian students, who reported lower levels of Course-Specific Self-Efficacy (CSSE) and PER, the students in the NAME department did not report lower levels of PER despite reporting lower CSSE. One reason for this could be the estimated coefficient for Asian students ($\beta = -0.17$, $p = 0.02$) was more than twice that of NAME students ($\beta = -0.07$, $p = 0.02$). Thus, given the marginally significant decrease in PER for Asian students ($\beta = -0.03$, $p = 0.07$), it is reasonable that any differences between NAME and ME students are not statistically significant.

Section Type

Finally, the qualitative differences in distributions by section type in Fig. 10.20 are revealed by the regression analysis illustrated in Fig. 10.21 to be nonsignificant.

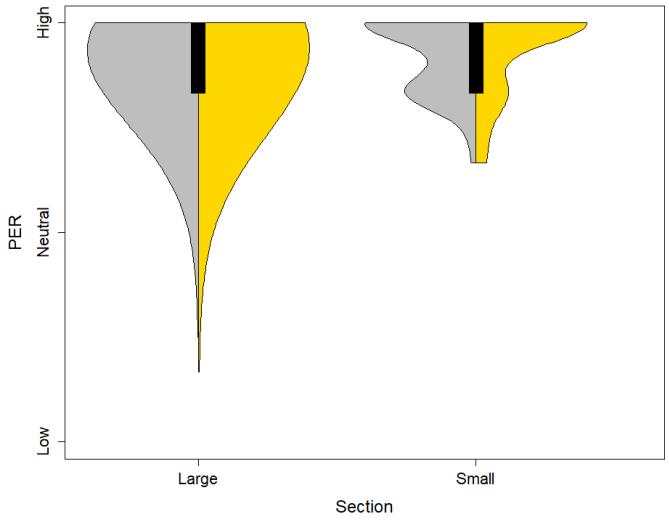


Figure 10.22: Split violin plots by section type of the Pre- and Post-score distributions (grey and gold, respectively) for PER.

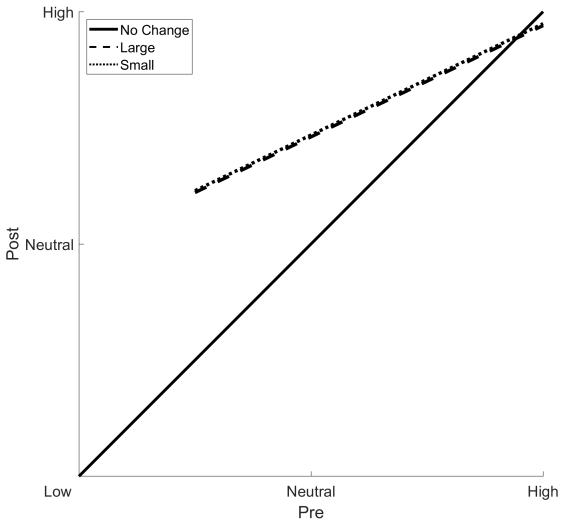


Figure 10.23: Results of the regression analyses by section type for PER. The black line denotes the baseline case that there was no change over the course of the semester. The other two lines distinguish section type.

Conclusions

Differences in Intention to Persist (PER) were largely the result of the three implementations of the active learning IMU interventions. The largest increases in PER, particularly for students with lower levels of PER at the start of the semester, arose with the greatest engagement with the iNewton IMU technology. Specifically, Level 2 (Prescribed Experiments) and Level 3 (Student Projects) were associated with statistically significant increases. Giving students the opportunity to apply ME240 concepts to a dynamic system of their choice appears to have improved PER, especially for students starting the semester with lower levels of PER who are arguably at the greatest risk of not persisting.

10.2.4 Study 4: Active Learning IMU Intervention and Student Affect

Study Background and Objective

The fourth and final study also leverages the findings from Study 1 to investigate differences in Student Affect (SA) between the three implementations of the active learning IMU intervention. This study is intended to provide additional insights into the students' experiences with the implementations of the intervention that involved the iNewton IMU technology, namely: 1) Demonstrations, 2) Prescribed Experiments, and 3) Student Projects. Since students historically respond positively towards hands-on experimentation (see, for example, O'Donovan et al. [80] or Self and Widmann [450]), the objective of this final study is to determine if students on average feel positively towards the active learning IMU intervention investigated in this work as well as to determine if there are any differences among the levels.

Methods

Students participated in one of the three intervention levels and were asked to report their Student Affect (SA) at the end of the semester. On a 7-point Likert scale, students reported the degree to which they agreed with the following statements from Table 10.1 (with corresponding item numbers):

13. The experiments improved my learning experience. (LEARN)

14. The experiments increased my interest in the subject matter. (INTEREST)

15. The experiments were engaging. (ENGAGE)

16. The experiments increased my confidence in running my own experiments. (CONFIDENCE)

Their answers were subsequently averaged and normalized to 1 such that a value of 0 corresponds to low SA (students strongly disagreed with all of the above items), a value of 0.5 corresponds to neutral SA, and a value of 1 corresponds to high SA (students strongly agreed with all of the above items).

The first objective is to determine if on average students feel positively towards the level of the intervention they experienced. A simple one-sample t-test comparing the mean SA to a hypothetical neutral SA mean is conducted for each intervention level. The second objective is to determine what (if any) differences exist across the intervention levels in SA, for which a one-way Analysis of Variance (ANOVA) with planned contrasts is conducted. The ANOVA is an omnibus test conducted to determine if at least one intervention level is significantly different from the others with respect to SA.

Assuming the ANOVA results are significant, the post hoc analyses conducted are planned contrasts, which are *a priori* comparisons of intervention level means that are expected to be different. For instance, Self and Widmann [450] reported that students in their study did not have a preference for demonstrations over hands-on experiments (or vice versa). Thus, the first planned contrast compares the average SA for Demonstrations and Prescribed Experiments. Assuming these two levels are not different from one another, the second planned contrast compares the combined average SA for Demonstrations and Prescribed Experiments against the average SA for Student Projects. Normality and homogeneity of variance are confirmed for the results.

Results and Discussion

Table 10.8 contains the results from the t-tests comparing the means of each intervention level's SA to a neutral SA (0.5). For all three intervention levels, the

results indicate that the null hypothesis that the intervention level SA mean is equal to 0.5 is rejected. Given the means, students generally had positive SA for all three intervention levels. It also appears that the SA for the Student Projects may be higher than the SA for Demonstrations or Prescribed Experiments.

Table 10.8: Summary of the one-sample t-tests. The Means and Standard Error (SE) for each intervention level is also reported.

Intervention Level	**N**	**Mean**	**SE**	**t-value**	**p-value**
Demonstrations	348	0.63	0.02	6.0	< 0.001
Prescribed Experiments	306	0.62	0.01	8.6	< 0.001
Student Projects	223	0.69	0.02	12.5	< 0.001

The results of the ANOVA ($F(2, 633) = 5.9$, $p < 0.01$) suggests at least one intervention level is different from the others. As expected given the similar means for Demonstrations and Prescribed Experiments in Table 10.8, the first planned contrast ($t(633) = 0.6$, $p = 0.56$) confirms that students do not prefer Demonstrations over Prescribed Experiments (or vice versa). This finding supports the results reported by Self and Widmann [450]. Additionally, the second planned contrast ($t(633) = 3.0$, $p < 0.01$) confirms that students generally prefer Student Projects to either Demonstrations or Prescribed Experiments.

In addition to the aggregated means, it is also worth considering each of the individual items constituting the SA construct. These results can potentially clarify why students seem to prefer Student Projects to the other two implementations of the intervention. Figure 10.24 illustrates the means for each item by intervention level. The error bars are two times the standard error, thus representing the 95% confidence intervals for the item's mean. Confidence intervals that do not overlap indicate a statistically significant difference between the two means. All three intervention levels scored highest and most consistently on Item # 13 ("The experiments improved my learning experience"). Demonstrations and Student Projects scored about the same as each other and slightly higher than Prescribed Experiments on Item #14 ("The experiments increased my interest in the subject matter") and Item #15

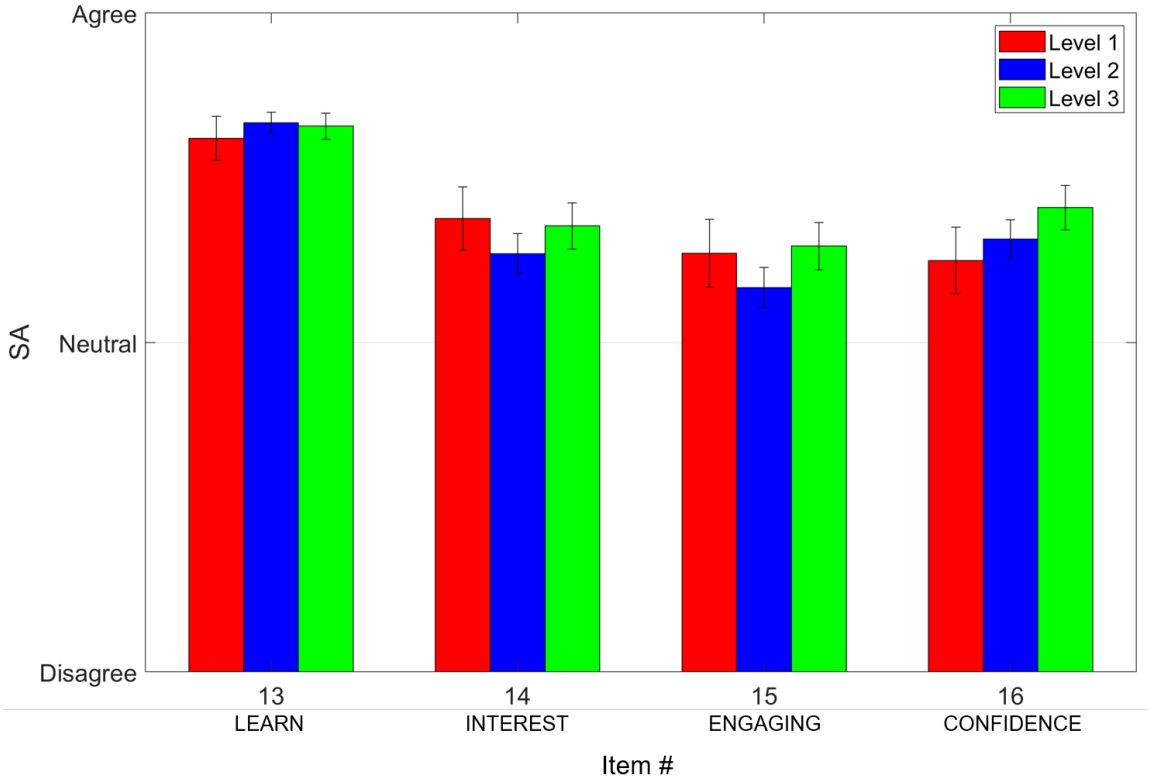


Figure 10.24: Student Affect by item number by intervention level: 1) Demonstrations, 2) Prescribed Experiments, and 3) Student Projects.

("The experiments were engaging"). The greatest difference between the three levels originated with Item #16 ("The experiments increased my confidence in running my own experiments").

Conclusions

In general, the students who participated in the various intervention levels all reported statistically significantly positive SA. Students did not seem to prefer Demonstrations over Prescribed Experiments or vice versa. However, the Student Projects version of the active learning IMU intervention yielded significantly higher SA than the other two levels, largely due to the item relating to the students' perceived confidence in conducting their own experiments.

10.3 Summary and Conclusions

This chapter investigated the validity and reliability of a modified version of the LAESE instrument and assessed the impacts of the active learning IMU interventions on self-efficacy (engineering and course-specific), intention to persist, and student affect. The first study confirmed the validity of the four constructs (i.e., they are distinguishably different from one another) via a confirmatory factor analysis, and it confirmed the reliability of the constructs via Cronbach's alpha. The modified version of the LAESE instrument was used in the remaining three studies with major conclusions summarized as follows.

The second study considered the effects of the intervention on Engineering Self-Efficacy (ESE) and Course-Specific Self-Efficacy (CSSE) with mixed results. Decreases in ESE over the term were associated with increasing levels of the intervention (i.e., increasing levels of cognitive engagement). However, the significant interaction between ESE Pre-score and intervention level indicated that students who entered ME240 with high ESE were largely unaffected by the intervention. For CSSE, there was a statistically significant increase for students who engaged with the Student Projects version (Level 3) of the intervention. Future work should consider how experiences with the iNewton IMU technology interact with experiences outside of the class. Other noteworthy results for CSSE included differences by ethnicity (including a significant interaction with Pre-score), a significant interaction between Pre-score and gender, and a significant difference between ME and NAME students.

The third study revealed that differences in Intention to Persist (PER) were largely associated with the active learning IMU intervention. The largest increases in PER, particularly for students with lower levels of PER at the start of the semester were associated with the Student Projects (Level 3) (i.e., the greatest engagement with the iNewton IMU technology). Specifically, Prescribed Experiments and Student Projects were associated with statistically significant increases in PER with the gains for Student Projects being greater than those for Prescribed Experiments. A marginal decrease in PER for Asian students was also detected.

The final study confirmed that students generally have a positive Student Af-

fect (SA) towards all of the interventions. There was no difference in SA between Demonstrations and Prescribed Experiments, but the SA for Student Projects were significantly higher.

CHAPTER 11

Part II: Summary, Contributions, and Conclusions

Summary and Contributions

Starting with Chapter 7, Part II of this dissertation provides ample evidence of the positive cognitive and non-cognitive impacts of active learning across disciplines drawing from the field of education research. In this research, active learning is systematically studied in a large undergraduate engineering dynamics course (ME240) in the form of an IMU intervention known as iNewton. It is hypothesized that the IMU intervention will elicit positive student responses in 1) conceptual understanding, 2) self-efficacy, and 3) intention to persist. These hypotheses are well-grounded in a theoretical foundation derived from the field of education research as reviewed in Chapter 7. The study design to test these hypotheses is outlined in Chapter 8, which provides detailed descriptions of the class setting, a pilot study, participant demographics, and each of the three levels of the active learning IMU intervention level (Demonstrations, Prescribed Experiments, and Student Projects).

Chapter 9 reports on the cognitive effects of the active learning IMU intervention via conceptual understanding as measured by the Dynamics Concept Inventory (DCI) [41, 42] in three studies. The first evaluates the authenticity of student performance on the DCI when administered in the low-stakes setting deployed in this research. The second investigates the reliability and validity of the DCI as an instrument for measuring conceptual understanding. The third assesses the impact of the three IMU intervention levels on conceptual understanding (as measured by the DCI). The

major contributions from these three studies are as follows. The first study confirms that conceptual understanding measured in a high-stakes setting (i.e., a midterm examination) is moderately correlated and predictive of conceptual understanding measured in a low-stakes setting (i.e., DCI scores from a voluntary survey) [394]. However, the regression models confirming the predictive relationship only explain a fraction (13%) of the variation in DCI performance. One reason is that the DCI covers many more concepts than what is covered in the high-stakes evaluation (a midterm examination). Another reason and contribution, revealed in the second study, is that 4 of the 29 items on the DCI were determined to be unreliable and thus subsequently removed from the statistical analysis conducted in the third study. Importantly, further investigation into the validity and reliability of the DCI revealed it may not have the ability to differentiate levels of conceptual understanding on a fine grain level. Future research should include revisions to the DCI to provide a more sensitive instrument to measure *changes* in conceptual understanding. The regression results in the third study reveal the active learning IMU intervention (all levels) had limited impact on student conceptual understanding, at least as measured by the DCI. However, significant differences in conceptual understanding by major and section size are also observed. Specifically, students taking ME240 as a technical elective tend to underperform on the DCI compared to those taking ME240 as a required course. Additionally, students enrolled in the small section offerings (fewer than 20 students) significantly outperform their peers enrolled in the large sections (60-120 students), which is consistent with other investigations into the effects of class size.

Chapter 10 begins with a study concerning the validity and reliability of the modified version of the Longitudinal Assessment of Engineering Self-Efficacy (LAESE) [46, 362] instrument designed to measure self-efficacy (engineering and course-specific), intention to persist, and student affect. This first study confirms the elements from various instruments are validly and reliably combined to simultaneously measure the four constructs constituting the instrument developed to measure these non-cognitive effects. Notably, this instrument is general enough for other researchers to use in their own contexts. A second study examines the effects of the active

learning IMU intervention on Engineering Self-Efficacy (ESE) and Course-Specific Self-Efficacy (CSSE) and exhibits mixed results. Decreases in ESE are associated with increasing levels of the intervention (i.e., increasing levels of cognitive engagement). However, the significant interaction between ESE Pre-score and intervention level indicates that students who enter ME240 with high ESE are also relatively unaffected by the intervention, which could be evidence of students who have developed resilient self-efficacy. For CSSE, there is a statistically significant increase for students who engaged with the Student Projects version (Level 3) of the intervention. Other significant results include differences in CSSE by ethnicity (including a significant interaction with Pre-score), the interaction between Pre-score and gender, and the difference between ME and NAME students. Given the differences in the results for Course-Specific Self-Efficacy and Engineering Self-Efficacy, future work should consider how experiences outside of the course interact with those related to the intervention. A third study reveals differences in Intention to Persist (PER) are significantly associated with the active learning IMU intervention. The largest increases in PER, particularly for students with lower PER at the start of the semester, relate to greatest engagement with the IMU technology. Specifically, Prescribed Experiments and Student Projects are associated with statistically significant increases in PER with the gains associated with Student Projects being the largest. A final (fourth) study confirms that students have a positive Student Affect (SA) towards the interventions. While there is no significant difference in SA between Demonstrations and Prescribed Experiments, there is a significant increase in SA associated with Student Projects.

Conclusions

This work systematically introduces and evaluates the effects of a very modest active learning IMU intervention within a traditionally taught (lecture only) engineering dynamics class. The implementation of the intervention takes into consideration two of the most commonly cited reasons for why faculty do not adopt active learning instructional methods, namely: 1) time (both in and out of the classroom) and 2) feasibility of implementing active learning in large classrooms. All levels of the interven-

tion occupy a modest fraction of the course (just two experiments) that students primarily engage with outside of the classroom. While Demonstrations and Prescribed Experiments require very little effort from the instructors, Student Projects necessitate more time outside of class to provide feedback. This time spent outside of class could be potentially offset in future iterations with peer review, which has the added benefit of an authentic educational opportunity to apply skills learned in the class as well. Furthermore, documentation for the experiments and findings to date are accessible to any instructors interested in implementing any version of the intervention in their engineering dynamics courses (`https://inewton.engin.umich.edu/`). In addition, this study design is somewhat unusual in that the levels of the intervention map onto three different levels of cognitive engagement outlined by Chi and Wylie [39]. Aside from the study conducted by Chi and Wylie that considers all four modes of engagement in the ICAP framework, the authors noted only two other studies conducted by Coleman et al. [466] and Gobert and Clement [467] that consider three modes of engagement in their study design (the same three considered in this work).

Overall, the active learning IMU intervention was successfully integrated into this large engineering dynamics course without significant in-class time commitment, with modest out-of-class time commitment, and with thorough documentation for faculty. While potential increases in conceptual understanding may not follow until higher doses are implemented as noted by Hake [258], the results for Course-Specific Self-Efficacy and Intention to Persist demonstrate the power of including even very modest levels of active learning into the curriculum. Further, the overarching idea of incorporating relatively inexpensive sensors into undergraduate courses to facilitate supplemental learning opportunities outside of the classroom can easily be extended to other courses. While the results for self-efficacy, intention to persist, and student affect may be transferable to other sophomore level courses, they may not transfer as strongly for upper level courses. For example, students are unlikely to switch majors their last year of their undergraduate career, meaning students' intention to persist in their major would likely remain relatively unaffected. However, this metric could be restructured to measure intention to persist in the *profession* after graduation. Finally, this work serves as evidence of the measurable improvements in

the undergraduate educational experience with a very minimal intervention. For faculty starting to adopt active learning instructional innovations into their classrooms, these interventions serve as a concrete starting point that have been shown to yield significant gains.

APPENDICES

APPENDIX A

Relationship Between Euler Angle Rates and Angular Velocity

Let R be a direction cosine matrix relating some body-fixed frame of reference (B) to some inertial frame of reference (I). R can be defined as a series of three subsequent Euler angle rotations (ϕ, θ, ψ) about body-fixed axis (X, Y, Z). The Euler direction cosine matrices are defined as

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & -\sin(\phi) \\ 0 & \sin(\phi) & \cos(\phi) \end{bmatrix} \tag{A.1}$$

$$R_y = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \tag{A.2}$$

$$R_z = \begin{bmatrix} \cos(\psi) & -\sin(\psi) & 0 \\ \sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{A.3}$$

Similarly, the Euler angle rates are defined in the inertial frame as

$$\dot{\phi}\hat{X} + \dot{\theta}\hat{Y} + \dot{\psi}\hat{Z} = \begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} \tag{A.4}$$

Therefore, the direction cosine matrix, R can therefore be defined as

$$R = R_x R_y R_z \tag{A.5}$$

The derivative of the rotation matrix is defined per [468]

$$\dot{R} = S({}^I\omega)R \tag{A.6}$$

where ${}^I\omega$ is the angular velocity of frame B relative to frame I (resolved in frame I) and the $S(\cdot)$ is the skew symmetric operator

$$S(\omega) = \begin{bmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{bmatrix} \tag{A.7}$$

Using the chain rule on Eq. A.5 and the relationship in Eq. A.6 yields

$$\begin{aligned} \dot{R} &= \dot{R}_x R_y R_z + R_x \dot{R}_y R_z + R_x R_y \dot{R}_z \\ &= S(\dot{\phi}\hat{X})R_x R_y R_z + R_x S(\dot{\theta}\hat{Y})R_y R_z + R_x R_y S(\dot{\psi}\hat{Z})R_z \\ &= S(\dot{\phi}\hat{X})R + R_x S(\dot{\theta}\hat{Y})R_x^T R + R_x R_y S(\dot{\psi}\hat{Z})(R_x R_y)^T R \\ &= S(\dot{\phi}\hat{X})R + S(\dot{\theta}R_x\hat{Y})R + S(\dot{\psi}R_x R_y\hat{Z})R \\ &= S(\dot{\phi}\hat{X} + \dot{\theta}R_x\hat{Y} + \dot{\psi}R_x R_y\hat{Z})R \end{aligned} \tag{A.8}$$

Equating Eqs. A.6 and A.6 yields

$$\begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & \sin\theta \\ 0 & \cos\phi & -\sin\phi\cos\theta \\ 0 & \sin\phi & \cos\phi\cos\theta \end{bmatrix} \begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} \tag{A.9}$$

or, similarly,

$$\begin{bmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} 1 & \frac{\sin\phi\sin\theta}{\cos\theta} & -\frac{\cos\phi\sin\theta}{\cos\theta} \\ 0 & \cos\phi & \sin\phi \\ 0 & -\frac{\sin\phi}{\cos\theta} & \frac{\cos\phi}{\cos\theta} \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix} \tag{A.10}$$

assuming that the matrix transforming Euler angle rates to angular velocity is invertible. The transformation matrix has singularities when $\theta = n\pi$ where $n = 0, \pm 1, \pm 2, ...$ (i.e., when $\cos\theta = 0$).

APPENDIX B

DCI Concept List

Below is a list of the 14 important and/or commonly misunderstood concepts identified in the DCI [42, 385].

1. Different points on a rigid body have different velocities and accelerations, which vary continuously.
2. If the net external force on a body is not zero, then the mass center must have an acceleration and it must be in the same direction as the force.
3. Angular velocities and angular accelerations are properties of the body as a whole and can vary with time.
4. Rigid bodies have both translational and rotational kinetic energy.
5. The angular momentum of a rigid body involves translational and rotational components and requires using some point as a reference.
6. Points on an object that is rolling without slip have velocities and accelerations that depend on the rolling without slip condition.
7. In general, the total mechanical energy is not conserved during an impact.

8. An object can have (a) nonzero acceleration and zero velocity or (b) nonzero velocity and no acceleration.

9. The inertia of a body affects its acceleration.

10. The direction of the friction force on a rolling body is not related in a fixed way to the direction of rolling.

11. A particle has acceleration when it is moving with a relative velocity on a rotating object.

12. An object moving in a curved path always has a normal component of acceleration.

13. The direction of the friction force between two objects depends on their relative velocity or their tendency for relative motion.

14. Newton's third law dictates that the interaction forces between two objects must be equal and opposite.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] D. Titterton and J.L. Weston. *Strapdown Inertial Navigation Technology*. Electromagnetics and Radar Series. Institution of Engineering and Technology, 2004. ISBN 978-0-86341-358-2.

[2] P.G. Savage. *Introduction to Strapdown Inertial Navigation Systems*. Strapdown Associates, 1991.

[3] J.C. Wilcox. A new algorithm for strapped-down inertial navigation. In *Proc. of IEEE Transactions on Aerospace and Electronic Systems*, pages 796–802, September 1967.

[4] C. Grubin. Attitude determiantion for a strapdown inertial system using the euler axis/angle and quaternion parameters. In *Proc. of AIAA Guidance and Control Conference*, pages No. 73–900, Key Biscayne, August 1973.

[5] N.M. Barbour. *Inertial Navigation Sensors*. Charles Stark Draper Lab Inc., Cambridge, UK, 2010.

[6] N. Yazdi, F. Ayazi, and K. Najafi. Micromachined inertial sensors. *Proceedings of the IEEE*, 86(8):1640–1659, Aug 1998.

[7] J.Z. Sasiadek. Sensor fusion. *Annual Reviews in Control*, 26(2):203–228, 2002.

[8] A. Filippeschi, N. Schmitz, M. Miezal, G. Bleser, E. Ruffaldi, and D. Stricker. Survey of motion tracking methods based on inertial sensors: A focus on upper limb human motion. *Sensors*, 17(1257), 2017.

[9] P. Eichelberger, M. Ferraro, U. Minder, T. Denton, A. Blasimann, F. Krause, and H. Baur. Analysis of accuracy in optical motion capture a protocol for laboratory setup evaluation. *Journal of Biomechanics*, 49(10):2085 – 2088, 2016.

[10] A.M. Aurand, J.S. Dufour, and W.S. Marras. Accuracy map of an optical motion capture system with 42 or 21 cameras in a large measurement volume. *Journal of Biomechanics*, 58:237 – 240, 2017.

[11] C. Wong, Z.Q. Zhang, B. Lo, and G.Z. Yang. Wearable sensing for solid biomechanics: A review. *IEEE Sensors Journal*, 15(5):2747–2760, 2015.

[12] T.D. Simon, C. Bublitz, and S.J. Hambidge. Emergency department visits among pediatric patients for sports-related injury: basic epidemiology and impact of race/ethnicity and insurance status. *Pediatric emergency care*, 22 (5):309–315, 2006.

[13] J. Chen, J. Kim, W. Shao, S.H. Schlecht, S.Y. Baek, A.K. Jones, T. Ahn, J.A. Ashton-Miller, M.M. Banaszak Holl, and E.M. Wojtys. An anterior cruciate ligament failure mechanism. *The American journal of sports medicine*, 47(9): 2067–2076, 2019.

[14] G.L. Hatfield, L. Cheryl C.L. Hubley-Kozey, A. Wilson, L. Janie, and M.J. Dunbar. The Effect of Total Knee Arthroplasty on Knee Joint Kinematics and Kinetics During Gait. *Journal of Arthroplasty*, 26(2):309–318, 2011.

[15] C.H. Wise. *Orthopaedic Manual Physical Therapy : From Art to Evidence*. Philadelphia, PA : F.A. Davis Company, 2015.

[16] D. Tack, A. Kelly, M. Richter, and J. Bray-Miner. Preliminary results of mc-leap testing of u.s. marine combat load order configurations. Technical Report ONR Contract Report No. N00014011-C-020, Quantico, VA: U.S. Marine Corps Systems Command, 2012.

[17] R.V. Vitali, S.M. Cain, S.P. Davidson, and N.C. Perkins. Human crawling performance and technique revealed by inertial measurement units. *Journal of Biomechanics*, 84:121–128, 2019.

[18] G. Wu and P.R. Cavanagh. ISB Recommendations for Standardization in the Reporting of Kinematic Data. *Journal of Biomechanics*, 28(10):1257–1261, 1995.

[19] T. Seel, J. Raisch, and T. Schauer. IMU-Based Joint Angle Measurement for Gait Analysis. *Sensors*, 14:6891–6909, 2014.

[20] R.V. Vitali, S.M. Cain, R.S. McGinnis, A.M. Zaferiou, L.V. Ojeda, S.P. Davidson, and N.C. Perkins. Method for Estimating Three-Dimensional Knee Rotations Using Two Inertial Measurement Units: Validation with a Coordinate Measurement Machine. *Sensors*, 17(1970), 2017.

[21] J.M. Cancio, R.A. Oliver, and K.E. Yancosek. Functional capacity evaluation-military: Program description and case series. *Military Medicine*, 182(1/2): e1658–e1664, 2017.

[22] H. Schermann, I. Karakis, O. Dolkart, E. Maman, A. Kadar, and O. Chechik. Olecranon bursitis in a military population: Epidemiology and evidence for prolonged morbidity in combat recruits. *Military Medicine*, 182(9/10):e1976–e1980, 2017.

[23] O. Schwartz, I. Malka, C.H. Olsen, I. Dudkiewicz, and T. Bader. Overuse injuries in the idf's combat training units: Rates, types, and mechanisms of injury. *Military Medicine*, 183(3/4):e196–e200, 2018.

[24] R.L. Jaworski, A. Jensen, B. Niederberger, R. Congalton, and K.R. Kelly. Changes in combat task performance under increasing loads in active duty marines. *Military Medicine*, 180(3):179–186, 2015.

[25] B. Larsen, K. Netto, D. Skovli, K. Vincs, S. Vu, and B. Aisbett. Body armor, performance, and physiology during repeated high-intensity work tasks. *Military Medicine*, 177(11):13081315, 2012.

[26] B. Larsen, K. Netto, and B. Aisbett. Task-specific effects of modular body armor. *Military Medicine*, 179(4):428–434, 2014.

[27] K.B. Mitchell, J.M. Batty, M.E. Coyne, L.L. DeSimone, and C.K. Bensel. Reliability analysis of time to complete the obstacle course portion of the load effects assessment program (leap). Technical Report NATICK/TR-17/002, Natick Soldier Research, Development, and Engineering Center (NSRDEC), Natick, MA, 2016.

[28] E.K. ONeal, J.H. Hornsby, and K.J. Kelleran. High-intensity tasks with external load in military applications: A review. *Military Medicine*, 179(9):950–954, 2014.

[29] M.J. Prince and R.M. Felder. Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education*, 95(2):123–138, 2006.

[30] M. Prince. Does active learning work? a review of the research. *Journal of Engineering Education*, 93(3):223–231, 2004.

[31] J.D. Bransford, A.L. Brown, and R.R. Cocking. *How people learn: Brain, mind, experience, and school*. Washington, D.C.: National Academy Press, 1999.

[32] J. Dewey. *Democracy and Education: An Introduction to the Philosophy of Education*. New York: Macmillan, 1916.

[33] N.E. Betz and R.S. Schifano. Evaluation of an intervention to increase realistic self-efficacy and interests in college women. *Journal of Vocational Behavior*, 56(1):552, 2000.

[34] M.W. Ohland, S.D. Sheppard, G. Lichtenstein, O. Eris, D. Chachra, and R.A. Layton. Persistence, engagement, and migration in engineering programs. *Journal of Engineering Education*, 97(3):259–278, 2013.

[35] J.E. Bean. Student attrition, intentions, and confidence: Interaction effects in a path model part i. the twenty-three variable model. In *American Education Research Association*, pages 1–48, Los Angeles, CA, April 1981.

[36] M.D. Svinicki. A Guidebook On Conceptual Frameworks For Research In Engineering Education, 2010.

[37] A. Bandura. *Social foundations of thought and action*. Englewood Cliffs, NJ, 1986.

[38] M.T.H. Chi. Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1:73105, 2009.

[39] M.T.H. Chi and R. Wylie. The icap framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4):219–243, 2014. doi: 10.1080/00461520.2014.965823.

[40] D. Montfort, S. Brown, and D. Pollock. An investigation of students' conceptual understanding in related sophomore to graduate-level engineering and mechanics courses. *Journal of Engineering Education*, 98(2):111–129, 04 2009.

[41] G.L. Gray, D. Evans, P. Cornwell, F. Costanzo, and B. Self. Toward a nationwide dynamics concept inventory assessment test. In *2003 ASEE Annual Conference and Exposition*, pages 1–12, Nashville, TN, June 2003.

[42] G.L. Gray, F. Costanzo, D. Evans, P. Cornwell, B. Self, and J.L. Lane. The dynamics concept inventory assessment test: A progress report and some results. In *2005 ASEE Annual Conference and Exposition*, pages 1–16, Portland, OR, June 2005.

[43] N. Jorion, B.D. Gane, K. James, L. Schroeder, L.V. DiBello, and J. Pellegrino. An analytic framework for evaluating the validity of concept inventory claims. *Journal of Engineering Education*, 104(4):454–496, 2015.

[44] A. Bandura. Self-efficacy mechanism in human agency. *American Psychology*, 37(2):122–147, 1982.

[45] A. Bandura. *Self-Efficacy: The Exercise of Control.* W H Freeman/Times Books/ Henry Holt & Co, New York, NY, US, 1997. ISBN 0-7167-2626-2 (Hardcover); 0-7167-2850-8 (Paperback).

[46] R.M. Marra, M. Schuurman, C. Moore, and B. Bogue. Women engineering students' self-efficacy beliefs: The longitudinal picture. In *2005 ASEE Annual Conference and Exposition*, pages 1–15, Portland, OR, June 2005.

[47] E. Arendt and R. Dick. Knee Injury Patterns Among Men and Women in Collegiate Basketball and Soccer. *The American Journal of Sports Medicine*, 1995. ISSN 0363-5465. doi: 10.1177/036354659502300611.

[48] C.M. Powers. The Influence of Abnormal Hip Mechanics on Knee Injury: A Biomechanical Perspective. *Journal of Orthopaedic & Sports Physical Therapy*, 40(2):42–51, 2010.

[49] E.M. Roos, H.P. Roos, S. Lohmander, C. Ekdahl, and B.D. Beynnon. Knee injury and osteoarthritis outcome score (koos)development of a self-administered outcome measure. *Journal Orthopaedic & Sports Physical Therapy*, page 8896, 1998.

[50] M.A. Davis, W.H. Ettinger, J.M. Neuhaus, S.A. Cho, and W.W. Hauck. The association of knee injury and obesity with unilateral and bilateral osteoarthritis of the knee. *American Journal of Epidemiology*, 130(2):278–288, 1989.

[51] J.Z. Edwards, K.A. Greene, R.S. Davis, M.W. Kovacik, D.A. Noe, and M.J. Askew. Measuring flexion in knee arthroplasty patients. *Journal of Arthroplasty*, 19(3):369–372, 2004.

[52] J. Favre, F. Luthi, B.M. Jolles, O. Siegrist, B. Najafi, and K. Aminian. A new ambulatory system for comparative evaluation of the three-dimensional knee kinematics, applied to anterior cruciate ligament injuries. *Knee Surg Sports Traumatol Arthrosc*, 14:592–604, 2006.

[53] R. Ferber, L.R. Osternig, M.H. Woollacott, N.J. Wasielewski, and J.H. Lee. Gait mechanics in chronic ACL deficiency and subsequent repair. *Clinical Biomechanics*, 17(4):274–285, 2002.

[54] J.R. Brinkmann and J. Perry. Rate and Range of Knee Motion During Ambulation Rate and Range of Knee Motion During Ambulation in Healthy and Arthritic Subjects. *American Journal of Physical Therapy*, 65:1055–1060, 1985.

[55] P.O. Almquist, C. Ekdahl, P.E. Isberg, and T. Fridén. Knee rotation in healthy individuals related to age and gender. *Journal of Orthopaedic Research*, 31(1): 23–28, 2013.

[56] X. Qu and J.C. Yeo. Effects of load carriage and fatigue on gait characteristics. *Journal of Biomechanics*, 44(7):1259–1263, 2011.

[57] S.A. Birrell and R.A. Haslam. The effect of load distribution within military load carriage systems on the kinetics of human gait. *Applied Ergonomics*, 41: 585–590, 2010.

[58] T.W.P. Huang and A.D. Kuo. Mechanics and energetics of load carriage during human walking. *The Journal of Experimental Biology*, 217:605–613, 2014.

[59] A. Silder, S.L. Delp, and T. Besier. Men and women adopt similar walking mechanics and muscle activation patterns during load carriage. *Journal of Biomechanics*, 46(14):2522–2528, 2013.

[60] G.M. Strathy, E.Y. Chao, and R.K. Laughman. Changes in knee function associated with treadmill ambulation. *Journal of Biomechanics*, 16(7), 1983.

[61] F. Alton, L. Baldey, S. Caplan, and M.C. Morrissey. A kinematic comparison of overground and treadmill walking. *Clinical Biomechanics*, 13(6):434 – 440, 1998.

[62] P.O. Riley, G. Paolini, U. Della Croce, K.W. Paylo, and D.C. Kerrigan. A kinematic and kinetic comparison of overground and treadmill walking in healthy subjects. *Gait and Posture*, 26(1):17 – 24, 2007.

[63] T. De Oliveira Sato, G.Å. Hansson, and H.J.C.G. Coury. Goniometer crosstalk compensation for knee joint applications. *Sensors (Switzerland)*, 10(11):9994–10005, 2010.

[64] S. Yabukami, H. Kikuchi, M. Yamaguchi, K.I. Arai, K. Takahashi, A. Itagaki, and N. Wako. Motion capture system of magnetic markers using three-axial magnetic field sensor. *IEEE Transactions on Magnetics*, 36(5):3646–3648, Sept 2000.

[65] A.D. Georgoulis, A. Papadonikolakis, C.D. Papageorgiou, A. Mitsou, and N. Stergiou. Three-Dimensional Tibiofemoral Kinematics of the Anterior Cruciate Ligament-Deficient and Reconstructed Knee during Walking <sup/>. *The American Journal of Sports Medicine*, 31(1):75–79, 2003.

[66] L.P. Maletsky, J. Sun, and N.A. Morton. Accuracy of an optical active-marker system to track the relative motion of rigid bodies. *Journal of Biomechanics*, 40(3):682 – 685, 2007.

[67] H. Zhou and H. Hu. Human motion tracking for rehabilitationa survey. *Biomedical Signal Processing and Control*, 3(1):1 – 18, 2008.

[68] S. Corazza, L. Mündermann, A.M. Chaudhari, T. Demattio, C. Cobelli, and T.P. Andriacchi. A markerless motion capture system to study musculoskeletal biomechanics: Visual hull and simulated annealing approach. *Annals of Biomedical Engineering*, 34(6):1019–1029, 2006.

[69] R. Poppe. Vision-based human motion analysis: An overview. *Computer Vision and Image Understanding*, 108(1):4 – 18, 2007. Special Issue on Vision for Human-Computer Interaction.

[70] H.J. Luinge, P.H. Veltink, and C.T.M. Baten. Ambulatory measurement of arm orientation. *Journal of Biomechanics*, 40(1):78–85, 2007.

[71] P. Müller, M.A. Bégin, T. Schauer, and T. Seel. Alignment-Free, Self-Calibrating Elbow Angles Measurement using Inertial Sensors. *IEEE Journal of Biomedical and Health Informatics*, 21(2):312 – 319, 2016.

[72] G. Cooper, I. Sheret, L. McMillian, K. Siliverdis, N. Sha, D. Hodgins, L. Kenney, and D. Howard. Inertial sensor-based knee flexion/extension angle estimation. *Journal of Biomechanics*, 42(16):2678–2685, 2009.

[73] E. Laidig, T. Schauer, and T. Seel. Exploiting kinematic constraints to compensate magnetic disturbances when calculating joint angles of approximate hinge joints from orientation estimates of inertial sensors. In *International Conference on Rehabilitation Robotics*, London, UK, July 2017.

[74] J. Favre, B.M. Jolles, R. Aissaoui, and K. Aminian. Ambulatory measurement of 3D knee joint angle. *Journal of Biomechanics*, 41:10291035, 2008.

[75] A. Brennan, J. Zhang K. Deluzio, and Q. Li. Quantification of inertial sensor-based 3D joint angle measurement accuracy using an instrumented gimbal. *Gait and Posture*, 34(3):320–323, 2011.

[76] R. Zhu and Z. Zhou. A real-time articulated human motion tracking using tri-axis inertial/magnetic sensors package. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 12(2):295–302, 2004.

[77] D. Roetenberg, H.J. Luinge, C.T.M. Baten, and P.H. Veltink. Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(3):395–405, 2005. doi: 10.1109/TNSRE.2005.847353.

[78] E. Bergamini, G. Ligorio, A. Summa, G. Vannozzi, A. Cappozzo, and A.M. Sabatini. Estimating orientation using magnetic and inertial sensors and different sensor fusion approaches: Accuracy assessment in manual and locomotion tasks. *Sensors (Switzerland)*, 14(10):18625–18649, 2014.

[79] T. Beravs, P. Reberšek, D. Novak, J. Podobnik, and M. Munih. Development and validation of a wearable inertial measurement system for use with lower limb exoskeletons. In *IEEE-RAS International Conference on Humanoid Robots*, pages 212–217, Bled, Slovenia, October 2011.

[80] K.J. O'Donovan, R. Kamnik, D.T. O'Keeffe, and G.M. Lyons. An inertial and magnetic sensor based technique for joint angle measurement. *Journal of Biomechanics*, 40(12):2604–2611, 2007.

[81] A. Cereatti, D. Trojaniello, and U. Della Croce. Accurately measuring human movement using magneto-inertial sensors: techniques and challenges. In *2015 IEEE International Symposium on Inertial Sensors and Systems (ISISS)*, pages 1–4, Hapuna Beach, HI, March 2015.

[82] M.D. Guillet. Design and test of a displacement workspace mapping station for articular joints, August 2006.

[83] R.M. Murray, Z. Li, and S.S. Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press: Boca Raton, FL, 1991.

[84] O.J. Woodman. An introduction to inertial navigation, 2007.

[85] S.O.H. Madgwick, A.J.L. Harrison, and R. Vaidyanathan. Estimation of imu and marg orientation using a gradient descent algorithm. In *2011 IEEE International Conference on Rehabilitation Robotics*, pages 1–7, Zurich, Switzerland, June 2011. doi: 10.1109/ICORR.2011.5975346.

[86] K. Feng, J. Li, X. Zhang, C. Shen, Y. Bi, T. Zheng, and J. Liu. A new quaternion-based kalman filter for real-time attitude estimation using the two-step geometrically-intuitive correction algorithm. *Sensors*, 17(9):2146, 2017.

[87] M. Raitoharju and R. Piché. On computational complexity reduction methods for kalman filter extensions. *arXiv*, 2015.

[88] A.L.C. Quigley. An approach to the control of divergence in kalman filter algorithms. *International Journal of Control*, 17(4):741–746, 1973.

[89] D. Gebre-Egziabher, G.H. Elkaim, J.D. Powell, and B.W. Parkinson. A non-linear, two-step estimation algorithm for calibrating solid-state strapdown magnetometers. In *2001 International Conference on Navigation Systems*, pages 28–30, St. Petersburg, Russia, May 2001.

[90] A.M. Sabatini. Quaternion-based extended kalman filter for determining orientation by inertial and magnetic sensing. *IEEE Transactions on Biomedical Engineering*, 53(7):1346–1356, 2006.

[91] V. Madyastha, V. Ravindra, S. Mallikarjunan, and A. Goyal. Extended kalman filter vs. error state kalman filter for aircraft attitude estimation. In *AIAA Guidance, Navigation, and Control Conference*, pages 1–23, Portland, OR, August 2011.

[92] S.J. Julier and J.K. Uhlmann. New extension of the kalman filter to nonlinear systems. In *Signal processing, sensor fusion, and target recognition VI*, volume 3068, pages 182–193, 1997.

[93] J.J. LaViola. A comparison of unscented and extended kalman filtering for estimating quaternion motion. In *Proceedings of the 2003 American Control Conference, 2003.*, volume 3, pages 2435–2440, June 2003. doi: 10.1109/ACC.2003.1243440.

[94] A. Yadav, D. Subedi, M.A. Lundeberg, and C.F. Bunting. Problem-based learning: Influence on students' learning in an electrical engineering course. *Journal of Engineering Education*, 100(2):253–280, 2011.

[95] R. Hartley, M.G. Jadidi, R.M. Eustice, and J.W. Grizzle. Contact-aided invariant extended kalman filtering for robot state estimation. *CoRR*, abs/1904.09251, 2019.

[96] S.I. Roumeliotis, G.S. Sukhatme, and G.A. Bekey. Circumventing dynamic modeling: evaluation of the error-state kalman filter applied to mobile robot localization. In *1999 IEEE International Conference on Robot Localization*, pages 1656–1663, Detroit, MI, May 1999.

[97] Joan Solà. Quaternion kinematics for the error-state kalman filter. *CoRR*, abs/1711.02508, 2017. URL `http://arxiv.org/abs/1711.02508`.

[98] N. Trawny and S.I. Roumeliotis. Indirect Kalman filter for 3D attitude estimation. Technical Report 2005-002, University of Minnesota, Department of Computer Science and Engineering, 2005.

[99] G. Welch and G. Bishop. An introduction to the kalman filter. Technical report, University of North Carolina, Department of Computer Science, 1995.

[100] F. Daum. Nonlinear filters: beyond the kalman filter. *IEEE Aerospace and Electronic Systems Magazine*, 20(8):57–69, 2005. doi: 10.1109/MAES.2005.1499276.

[101] M.I. Ribeiro. Kalman and extended kalman filters: Concept, derivation and properties. Technical report, Institute for Systems and Robotics, 2004.

[102] P.C. Mahalanobis. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, 2:49–55, 1936.

[103] N.K. Shah and P.J. Gemperline. Combination of the Mahalanobis distance and residual variance pattern recognition techniques for classification of near-infrared reflectance spectra. *Analytical Chemistry*, 62(5):465–470, 1990.

[104] U. Della Croce, A. Leardini, L. Chiari, and A. Cappozzo. Human movement analysis using stereophotogrammetry: Part 4: assessment of anatomical landmark misplacement and its effects on joint kinematics. *Gait and Posture*, 21 (2):226 – 237, 2005.

[105] A. Brennan, J. Zhang, K. Deluzio, and Q. Li. Quantification of inertial sensor-based 3D joint angle measurement accuracy using an instrumented gimbal. *Gait and Posture*, 34(3):320–323, 2011.

[106] Q. Li and J.T. Zhang. Post-trial anatomical frame alignment procedure for comparison of 3d joint angle measurement from magnetic/inertial measurement units and camera-based systems. *Physiological Measurement*, 35(11): 2255–2268, 2014.

[107] K. Li, L. Zheng, S. Tashman, and X. Zhang. The inaccuracy of surface-measured model-derived tibiofemoral kinematics. *Journal of biomechanics*, 45 (15):2719–23, 2012.

[108] D.R. Hume, V. Kefala, M.D. Harris, and K.B. Shelburne. Comparison of marker-based and stereo radiography knee kinematics in activities of daily living. *Annals of Biomedical Engineering*, Jun 2018. doi: 10.1007/s10439-018-2068-9. URL `https://doi.org/10.1007/s10439-018-2068-9`.

[109] R. Stagni, S. Fantozzi, A. Cappello, and A. Leardini. Quantification of soft tissue artefact in motion analysis by combining 3d fluoroscopy and stereophotogrammetry: a study on two subjects. *Clinical Biomechanics*, 20(3):320 – 329, 2005.

[110] M. Akbarshahi, A.G. Schache, J.W. Fernandez, R. Baker, S. Banks, and M.G. Pandy. Non-invasive assessment of soft-tissue artifact and its effect on knee joint kinematics during functional activity. *Journal of Biomechanics*, 43(7): 1292 – 1301, 2010.

[111] N.M. Fiorentino, P.R. Atkins, M.J. Kutschke, J.M. Goebel, K.B. Foreman, and A.E. Anderson. Soft tissue artifact causes significant errors in the calculation of joint angles and range of motion at the hip. *Gait & posture*, 55:184–190, 2017.

[112] E.S. Grood and W.J. Suntay. A joint coordinate system for the clinical description of three-dimensional motions: Application to the knee. *ASME Journal of Biomechanical Engineering*, 105(2):136–144, 1983.

[113] G. Wu, S. Siegler, P. Allard, C. Kirtley, A. Leardini, D. Rosenbaum, M. Whittle, D.D. DLima, L. Cristofolini, H. Witte, O. Schmid, and I. Stokes. Isb recommendation on definitions of joint coordinate system of various joints for

the reporting of human joint motionpart i: ankle, hip, and spine. *Journal of Biomechanics*, 35(4):543 – 548, 2002.

[114] G. Wu, F.C.T. van der Helm, H.E.J. Veeger, M. Makhsous, P. Van Roy, C. Anglin, J. Nagels, A.R. Karduna, K. McQuade, X. Wang, F.W. Werner, and B. Buchholz. Isb recommendation on definitions of joint coordinate systems of various joints for the reporting of human joint motionpart ii: shoulder, elbow, wrist and hand. *Journal of Biomechanics*, 38(5):981 – 992, 2005.

[115] A.G. Cutti, A. Giovanardi, L. Rocchi, A. Davalli, and R. Sacchetti. Ambulatory measurement of shoulder and elbow kinematics through inertial and magnetic sensors. *Medical and Biological Engineering and Computing*, 46(2):169–178, 2008. doi: 10.1007/s11517-007-0296-5.

[116] J.H.M. Bergmann, R.E. Mayagoitia, and I.C.H. Smith. A portable system for collecting anatomical joint angles during stair ascent: A comparison with an optical tracking device. *Dynamic Medicine*, 8(3), 2009. doi: 10.1186/1476-5918-8-3.

[117] J. Favre, R. Aissaoui, B.M. Jolles, J.A. de Guise, and K. Aminian. Functional calibration procedure for 3D knee joint angle description using inertial sensors. *Journal of Biomechanics*, 42(14):2330–2335, 2009. doi: 10.1016/j.jbiomech.2009.06.025.

[118] R. Takeda, S. Tadano, A. Natorigawa, M. Todoh, and S. Yoshinari. Gait posture estimation using wearable acceleration and gyro sensors. *Journal of Biomechanics*, 42(15):2486–2494, 2009.

[119] A.G. Cutti, A. Ferrari, P. Garofalo, M. Raggi, A. Cappello, and A. Ferrari. 'Outwalk': A protocol for clinical gait analysis based on inertial and magnetic sensors. *Medical and Biological Engineering and Computing*, 48(1):17–25, 2010. doi: 10.1007/s11517-009-0545-x.

[120] W.H.K. de Vries, H.E.J. Veeger, A.G. Cutti, C. Baten, and F.C.T. van der Helm. Functionally interpretable local coordinate systems for the upper extremity using inertial & magnetic measurement systems. *Journal of Biomechanics*, 43(10):1983–1988, 2010. doi: 10.1016/j.jbiomech.2010.03.007.

[121] Z.Q. Zhang and J.K. Wu. A novel hierarchical information fusion method for three-dimensional upper limb motion estimation. *IEEE Transactions on Instrumentation and Measurement*, 60(11):3709–3719, 2011. doi: 10.1109/TIM.2011.2135070.

[122] J. Chardonnens, J. Favre, B. Le Callennec, F. Cuendet, G. Gremion, and K. Aminian. Automatic measurement of key ski jumping phases and temporal events with a wearable system. *Journal of Sports Sciences*, 30(1):53–61, 2012. doi: 10.1080/02640414.2011.624538.

[123] L.I.F. Penning, N.A. Guldemond, R.A. de Bie, and G.H.I.M. Walenkamp. Reproducibility of a 3-dimensional gyroscope in measuring shoulder anteflexion and abduction. *BMC Musculoskeletal Disorders*, 13(135), 2012.

[124] Z.Q. Zhang, L.Y. Ji, Z.P. Huang, and J.K. Wu. Adaptive information fusion for human upper limb movement estimation. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, 42(5):1100–1108, 2012. doi: 10.1109/TSMCA.2012.2189876.

[125] J. Chardonnens, J. Favre, F. Cuendet, G. Gremion, and K. Aminian. A system to measure the kinematics during the entire ski jump sequence using inertial sensors. *Journal of Biomechanics*, 46(1):56–62, 2013. doi: 10.1016/j.jbiomech.2012.10.005.

[126] B.F. El-Zayat, T. Efe, A. Heidrich, R. Anetsmann, N. Timmesfeld, S. Fuchs-Winkelmann, and M.D. Schofer. Objective assessment, repeatability, and agreement of shoulder ROM with a 3D gyroscope. *BMC Musculoskeletal Disorders*, 14(72):1–7, 2013. doi: 10.1186/1471-2474-14-72.

[127] Á. Gil-Agudo, A. de los Reyes-Guzmán, I. Dimbwadyo-Terrer, B. Peñasco-Martín, A. Bernal-Sahún, P. López-Monteagudo, A. del Ama-Espinosa, and J.L. Pons. A novel motion tracking system for evaluation of functional rehabilitation of the upper limbs. *Neural Regeneration Research*, 8(19):1773–1782, 2013. doi: 10.3969/j.issn.1673-5374.2013.19.005.

[128] S. Kim and M.A. Nussbaum. Performance evaluation of a wearable inertial motion capture system for capturing physical exposures during manual material handling tasks. *Ergonomics*, 56(2):314–326, 2013. doi: 10.1080/00140139.2012.742932.

[129] J.N. Kim, M.H. Ryu, H.R. Choi, and Y.S. Yang. Anatomy calibration of inertial measurement unit using a principle component analysis. *International Journal of Bio-Science and Bio-Technology*, 5(6):181–189, 2013. doi: 10.14257/ijbsbt.2013.5.6.19.

[130] A. Laudanski, B. Brouwer, and Q. Li. Measurement of Lower Limb Joint Kinematics using Inertial Sensors During Stair Ascent and Descent in Healthy Older Adults and Stroke Survivors. *Journal of Healthcare Engineering*, 4(4): 555–576, 2013. doi: 10.1260/2040-2295.4.4.555.

[131] F. Öhberg, R. Lundström, and H. Grip. Comparative analysis of different adaptive filters for tracking lower segments of a human body using inertial motion sensors. *Measurement Science and Technology*, 24(8), 2013. doi: 10.1088/0957-0233/24/8/085703.

[132] S. Tadano, R. Takeda, and H. Miyagawa. Three dimensional gait analysis using wearable acceleration and gyro sensors based on quaternion calculations. *Sensors*, 13(7):9321–9343, 2013. doi: 10.3390/s130709321.

[133] J.T. Zhang, A.C. Novak, B. Brouwer, and Q. Li. Concurrent validation of Xsens MVN measurement of lower limb joint angular kinematics. *Physiological Measurement*, 34(8), 2013. doi: 10.1088/0967-3334/34/8/N63.

[134] B. Bouvier, A. Savescu, S. Duprey, and R. Dumas. Benefits of functional calibration for estimating elbow joint angles using magneto-inertial sensors: preliminary results. *Computer Methods in Biomechanics and Biomedical Engineering*, 17:108–109, 2014. doi: 10.1080/10255842.2014.931444.

[135] J. Cockcroft, J.H. Muller, and C. Scheffer. A novel complimentary filter for tracking hip angles during cycling using wireless inertial sensors and dynamic acceleration estimation. *IEEE Sensors Journal*, 14(8):2864–2871, 2014. doi: 10.1109/JSEN.2014.2318897.

[136] A. Leardini, G. Lullini, S. Giannini, L. Berti, M. Ortolani, and P. Caravaggi. Validation of the angular measurements of a new inertial-measurement-unit based rehabilitation system: comparison with state-of-the-art gait analysis. *Journal of Neuroengineering and Rehabilitation*, 11(1):136, 2014.

[137] E. Palermo, S. Rossi, F. Marini, F. Patanè, and P. Cappa. Experimental evaluation of accuracy and repeatability of a novel body-to-sensor calibration procedure for inertial sensor-based gait analysis. *Measurement: Journal of the International Measurement Confederation*, 52(1):145–155, 2014. doi: 10.1016/j.measurement.2014.03.004.

[138] L. Ricci, D. Formica, L. Sparaci, F. Romana Lasorsa, F. Taffoni, E. Tamilia, and E. Guglielmelli. A new calibration methodology for thorax and upper

limbs motion capture in children using magneto and inertial sensors. *Sensors*, 14(1):1057–1072, 2014. doi: 10.3390/s140101057.

[139] B. Bouvier, S. Duprey, L. Claudon, R. Dumas, and A. Savescu. Upper limb kinematics using inertial and magnetic sensors: Comparison of sensor-to-segment calibrations. *Sensors*, 15(8):18813–18833, 2015. doi: 10.3390/s150818813.

[140] S. Fantozzi, A. Giovanardi, D. Borra, and G. Gatta. Gait kinematic analysis in water using wearable inertial magnetic sensors. *PLOS ONE*, 10(9), 2015. doi: 10.1371/journal.pone.0138105.

[141] T. Khurelbaatar, K. Kim, S. Lee, and Y. Kim. Consistent accuracy in whole-body joint kinetics during gait using wearable inertial motion sensors and in-shoe pressure sensors. *Gait and Posture*, 42(1):65–69, 2015. doi: 10.1016/j.gaitpost.2015.04.007.

[142] G. Logar and M. Munih. Estimation of joint forces and moments for the in-run and take-off in ski jumping based on measurements with wearable inertial sensors. *Sensors*, 15:11258–11276, 2015. doi: 10.3390/s150511258.

[143] J.C. van den Noort, S.H. Wiertsema, K.M.C Hekman, C.P. Schönhuth, J. Dekker, and J. Harlaar. Measurement of scapular dyskinesis using wireless inertial and magnetic sensors: Importance of scapula calibration. *Journal of Biomechanics*, 48(12):3460–3468, 2015. doi: 10.1016/j.jbiomech.2015.05.036.

[144] D. Álvarez, J.C. Alvarez, R.C. González, and A.M. López. Upper limb joint angle measurement in occupational health. *Computer Methods in Biomechanics and Biomedical Engineering*, 19(2):159–170, 2016. doi: 10.1080/10255842.2014.997718.

[145] P. Ertzgaard, F. Öhberg, B. Gerdle, and H. Grip. A new way of assessing arm function in activity using kinematic Exposure Variation Analysis and portable inertial sensors - A validity study. *Manual Therapy*, 21:241–249, 2016. doi: 10.1016/j.math.2015.09.004.

[146] S. Fantozzi, A. Giovanardi, F.A. Magalhães, R. Di Michele, M. Cortesi, and G. Gatta. Assessment of three-dimensional joint kinematics of the upper limb during simulated swimming using wearable inertial-magnetic measurement units. *Journal of Sports Sciences*, 34(11):1073–1080, 2016. doi: 10.1080/02640414.2015.1088659.

[147] W. Kong, S. Sessa, M. Zecca, and A. Takanishi. Anatomical calibration through post-processing of standard motion tests data. *Sensors*, 16(12), 2016. doi: 10.3390/s16122011.

[148] G. Li, T. Liu, J. Yi, H. Wang, J. Li, and Y. Inoue. The lower limbs kinematics analysis by wearable sensor shoes. *IEEE Sensors Journal*, 16(8):2627–2638, 2016. doi: 10.1109/JSEN.2016.2515101.

[149] J. Reenalda, E. Maartens, L. Homan, and J.H. Buurke. Continuous three dimensional analysis of running mechanics during a marathon by means of inertial magnetic measurement units to objectify changes in running mechanics. *Journal of Biomechanics*, 49(14):3362–3367, 2016. doi: 10.1016/j.jbiomech.2016.08.032.

[150] J.C. Van Den Noort, H.G. Kortier, N. Van Beek, D.H.E.J Veeger, and P.H. Veltink. Measuring 3D hand and finger kinematics - A comparison between inertial sensing and an opto-electronic marker system. *PLOS ONE*, 11(11), 2016. doi: 10.1371/journal.pone.0164889.

[151] L.S. Vargas-Valencia, A. Elias, E. Rocon, T. Bastos-Filho, and A. Frizera. An IMU-to-Body Alignment Method Applied to Human Gait Analysis. *Sensors*, 16(2090), 2016.

[152] B. Fasel, J. Spörri, P. Schütz, S. Lorenzetti, and K. Aminian. An inertial sensor-based method for estimating the athlete's relative joint center positions and center of mass kinematics in alpine ski racing. *Frontiers in Physiology*, 8 (850), 2017. doi: 10.3389/fphys.2017.00850.

[153] B. Fasel, J. Spörri, P. Schütz, S. Lorenzetti, and K. Aminian. Validation of functional calibration and strap-down joint drift correction for computing 3D joint angles of knee, hip, and trunk in alpine skiing. *PLOS ONE*, 12(7), 2017. doi: 10.1371/journal.pone.0181446.

[154] C. Nüesch, E. Roos, G. Pagenstert, and A. Mündermann. Measuring joint kinematics of treadmill walking and running: Comparison between an inertial sensor based system and a camera-based system. *Journal of Biomechanics*, 57: 32–38, 2017. doi: 10.1016/j.jbiomech.2017.03.015.

[155] M. Kim and D. Lee. Wearable inertial sensor based parametric calibration of lower-limb kinematics. *Sensors and Actuators A: Physical*, 265:280–296, 2017. doi: 10.1016/j.sna.2017.07.017.

[156] G. Ligorio, D. Zanotto, A.M. Sabatini, and S.K. Agrawal. A novel functional calibration method for real-time elbow joint angles estimation with magnetic-inertial sensors. *Journal of Biomechanics*, 54:106–110, 2017. doi: 10.1016/j.jbiomech.2017.01.024.

[157] A.L. Mangia, C. Cortesi, S. Fantozzi, A. Giovanardi, D. Borra, and G. Gatta. The use of IMMUs in a water environment: Instrument validation and application of 3D multi-body kinematic analysis in medicine and sport. *Sensors*, 17 (4), 2017. doi: 10.3390/s17040927.

[158] X. Robert-Lachaine, H. Mecheri, C. Larue, and A. Plamondon. Accuracy and repeatability of single-pose calibration of inertial measurement units for whole-body motion analysis. *Gait and Posture*, 54:80–86, 2017. doi: 10.1016/j.gaitpost.2017.02.029.

[159] X. Robert-Lachaine, H. Mecheri, C. Larue, and A. Plamondon. Validation of inertial measurement units with an optoelectronic system for whole-body motion analysis. *Medical and Biological Engineering and Computing*, 55(4): 609–619, 2017. doi: 10.1007/s11517-016-1537-2.

[160] M. Al-Amri, K. Nicholas, K. Button, V. Sparkes, L. Sheeran, and J.L. Davies. Inertial measurement units for clinical movement analysis: Reliability and concurrent validity. *Sensors*, 18(3), 2018. doi: 10.3390/s18030719.

[161] S. Blair, G. Duthie, S. Robertson, W. Hopkins, and K. Ball. Concurrent validation of an inertial measurement system to quantify kicking biomechanics in four football codes. *Journal of Biomechanics*, 73:24–32, 2018. doi: 10.1016/j.jbiomech.2018.03.031.

[162] B. Fasel, J. Sporri, J. Chardonnens, J. Kroll, E. Muller, and K. Aminian. Joint Inertial Sensor Orientation Drift Reduction for Highly Dynamic Movements. *IEEE Journal of Biomedical and Health Informatics*, 22(1):77–86, 2018. doi: 10.1109/JBHI.2017.2659758.

[163] M. Makikawa, S. Kurata, Y. Higa, Y. Araki, and R. Tokue. Ambulatory Monitoring of Behavior in Daily Life by Accelerometers Set at Both-Near-Sides of the Joint. *Studies in Health Technology and Informatics*, 84:840–843, 2001. doi: 10.3233/978-1-60750-928-8-840.

[164] Y. Ohtaki, K. Sagawa, and H. Inooka. A Method for Gain Analysis in a Daily Living Environment by Body-Mounted Instruments. *JSME International*

Journal Series: Mechanical Systems, Machine Elements and Manufacturing, 44 (4):1125–1132, 2001. doi: 10.1299/jsmec.44.1125.

[165] R. Williamson and B.J. Andrews. Detecting absolute human knee angle and angular velocity using accelerometers and rate gyroscopes. *Medical and Biological Engineering and Computing*, 39(3):294–302, 2001.

[166] H. Dejnabadi, B.M. Jolles, and K. Aminian. A new approach to accurate measurement of uniaxial joint angles based on a combination of accelerometers and gyroscopes. *IEEE Transactions on Biomedical Engineering*, 52(8):1478–1484, 2005. doi: 10.1109/TBME.2005.851475.

[167] H. Dejnabadi, B.M. Jolles, E. Casanova, P. Fua, and K. Aminian. Estimation and visualization of sagittal kinematics of lower limbs orientation using body-fixed sensors. *IEEE Transactions on Biomedical Engineering*, 53(7):1385–1393, 2006. doi: 10.1109/TBME.2006.873678.

[168] A. Findlow, J.Y. Goulermas, C. Nester, D. Howard, and L.P.J. Kenney. Predicting lower limb joint kinematics using wearable motion sensors. *Gait and Posture*, 28(1):120–126, 2008. doi: 10.1016/j.gaitpost.2007.11.001.

[169] R. Zheng, T. Liu, Y. Inoue, K. Shibata, and K. Liu. Kinetics Analysis of Ankle, Knee and Hip Joints Using a Wearable Sensor System. *Journal of Biomechanical Science and Engineering*, 3(3):343–355, 2008. doi: 10.1299/jbse.3.343.

[170] R. Takeda, S. Tadano, M. Todoh, M. Morikawa, M. Nakayasu, and S. Yoshinari. Gait analysis using gravitational acceleration measured by wearable sensors. *Journal of Biomechanics*, 42(3):223–233, 2009. doi: 10.1016/j.jbiomech.2008.10.027.

[171] A. Krüger and J. Edelmann-Nusser. Application of a full body inertial measurement system in alpine skiing: a comparison with an optical video based system. *Journal of Applied Biomechanics*, 26(4):516–21, 2010.

[172] K. Liu, Y. Inoue, and K. Shibata. Physical-Sensor and Virtual-Sensor Based Method for Estimation of Lower Limb Gait Posture Using Accelerometers and Gyroscopes. *Journal of Biomechanical Science and Engineering*, 5(4):472–483, 2010. doi: 10.1299/jbse.5.472.

[173] R. Pérez, Ú. Costa, M. Torrent, J. Solana, E. Opisso, C. Cáceres, J.M. Tormos, J. Medina, and E.J. Gómez. Upper limb portable motion analysis system based on inertial technology for neurorehabilitation purposes. *Sensors*, 10(12):10733–10751, 2010. doi: 10.3390/s101210733.

[174] M.D. Djurić-Jovičić, N.S. Jovičić, and D.B. Popović. Kinematics of gait: New method for angle estimation based on accelerometers. *Sensors*, 11(11):10571–10585, 2011. doi: 10.3390/s111110571.

[175] L. Kun, Y. Inoue, K. Shibata, and C. Enguo. Ambulatory estimation of knee-joint kinematics in anatomical coordinate system using accelerometers and magnetometers. *IEEE Transactions on Biomedical Engineering*, 58(2):435–442, 2011. doi: 10.1109/TBME.2010.2089454.

[176] W.D. Paulis, H.L.D. Horemans, B.S. Brouwer, and H.J. Stam. Excellent test-retest and inter-rater reliability for Tardieu Scale measurements with inertial sensors in elbow flexors of stroke patients. *Gait and Posture*, 33(2):185–189, 2011. doi: 10.1016/j.gaitpost.2010.10.094.

[177] H. Saito and T. Watanabe. Kalman-filtering-based joint angle measurement with wireless wearable sensor system for simplified gait analysis. *IEICE Transactions on Information and Systems*, 94(8):1716–1720, 2011.

[178] T. Watanabe, H. Saito, E. Koike, and K. Nitta. A Preliminary Test of Measurement of Joint Angles and Stride Length with Wireless Inertial Sensors for Wearable Gait Evaluation System. *Computational Intelligence and Neuroscience*, 2011(6):1–12, 2011. doi: 10.1155/2011/975193.

[179] M.D. Djurić-Jovičić, N.S. Jovičić, D.B. Popović, and A.R. Djordjević. Nonlinear optimization for drift removal in estimation of gait kinematics based on accelerometers. *Journal of Biomechanics*, 45(16):2849–2854, 2012. doi: 10.1016/j.jbiomech.2012.08.028.

[180] M. El-Gohary and J. McNames. Shoulder and elbow joint angle tracking with inertial sensors. *IEEE Transactions on Biomedical Engineering*, 59(9):2635–2641, 2012. doi: 10.1109/TBME.2012.2208750.

[181] Y. Guo, D. Wu, G. Liu, G. Zhao, B. Huang, and L. Wang. A Low-Cost Body Inertial-Sensing Network for Practical Gait Discrimination of Hemiplegia Patients. *Telemedicine and e-Health*, 18(10):748–754, 2012. doi: 10.1089/tmj.2012.0014.

[182] J.F.S. Lin and D. Kulić. Human pose recovery using wireless inertial measurement units. *Physiological Measurement*, 33(12):2099–2115, 2012. doi: 10.1088/0967-3334/33/12/2099.

[183] M. Ockendon and R. Gilbert. Validation of a Novel Smartphone Accelerometer-Based Knee Goniometer. *Journal of Knee Surgery*, 25(04):341–346, 2012. doi: 10.1055/s-0031-1299669.

[184] S.M.N. Arosha Senanayake, O.A. Malik, P.M. Iskandar, and D. Zaheer. Assessing post-anterior cruciate ligament reconstruction ambulation using wireless wearable integrated sensors. *Journal of Medical Engineering and Technology*, 37(8):498–510, 2013. doi: 10.3109/03091902.2013.837529.

[185] A. Caroselli, F. Bagalà, and A. Cappello. Quasi-real time estimation of angular kinematics using single-axis accelerometers. *Sensors*, 13(1):918–937, 2013. doi: 10.3390/s130100918.

[186] Y. Guo, G. Zhao, Q. Liu, Z. Mei, K. Ivanov, and L. Wang. Balance and knee extensibility evaluation of hemiplegic gait using an inertial body sensor network. *BioMedical Engineering Online*, 12(1), 2013. doi: 10.1186/1475-925X-12-83.

[187] F. Martínez-Solís, A. Claudio-Sánchez, J.M. Rodríguez-Lelis, S. Vergara-Limon, V. Olivares-Peregrino, and M. Vargas-Treviño. A portable system with sample rate of 250 Hz for characterization of knee and hip angles in the sagittal plane during gait. *BioMedical Engineering Online*, 13(1), 2014. doi: 10.1186/1475-925X-13-34.

[188] S. Šlajpah, R. Kamnik, and M. Munih. Kinematics based sensory fusion for wearable motion assessment in human walking. *Computer Methods and Programs in Biomedicine*, 116(2):131–144, 2014. doi: 10.1016/j.cmpb.2013.11.012.

[189] R. Takeda, G. Lisco, T. Fujisawa, L. Gastaldi, H. Tohyama, and S. Tadano. Drift removal for improving the accuracy of gait parameters using wearable sensor systems. *Sensors*, 14(12):23230–23247, 2014. doi: 10.3390/s141223230.

[190] K.H. Chen, P.C. Chen, K.C. Liu, and C.T. Chan. Wearable sensor-based rehabilitation exercise assessment for knee osteoarthritis. *Sensors*, 15(2):4193–4211, 2015. doi: 10.3390/s150204193.

[191] T. Jaysrichai, A. Suputtitada, and W. Khovidhungij. Mobile sensor application for kinematic detection of the knees. *Annals of Rehabilitation Medicine*, 39(4): 599–608, 2015. doi: 10.5535/arm.2015.39.4.599.

[192] J. Ma, H. Kharboutly, A. Benali, F. Benamar, and M. Bouzit. Joint angle estimation with accelerometers for dynamic postural analysis. *Journal of Biomechanics*, 48(13):3616–3624, 2015. doi: 10.1016/j.jbiomech.2015.08.008.

[193] P. Picerno, V. Viero, M. Donati, T. Triossi, V. Tancredi, and G. Melchiorri. Ambulatory assessment of shoulder abduction strength curve using a single wearable inertial sensor. *Journal of Rehabilitation Research and Development*, 52(2):171–180, 2015. doi: 10.1682/jrrd.2014.06.0146.

[194] A. Ahmadi, F. Destelle, L. Unzueta, D.S. Monaghan, M.T. Linaza, K. Moran, and N.E. O'Connor. 3D human gait reconstruction and monitoring using body-worn inertial sensors and kinematic modeling. *IEEE Sensors Journal*, 16(24): 8823–8831, 2016. doi: 10.1109/JSEN.2016.2593011.

[195] B. Kirking, M. El-Gohary, and Y. Kwon. The feasibility of shoulder motion tracking during activities of daily living using inertial measurement units. *Gait and Posture*, 49:47–53, 2016. doi: 10.1016/j.gaitpost.2016.06.008.

[196] J. Kodama and T. Watanabe. Examination of inertial sensor-based estimation methods of lower limb joint moments and ground reaction force: Results for squat and sit-to-stand movements in the sagittal plane. *Sensors*, 16(8), 2016. doi: 10.3390/s16081209.

[197] E.B. Mazomenos, D. Biswas, A. Cranny, A. Rajan, K. Maharatna, J. Achner, J. Klemke, M. Jöbges, S. Ortmann, and P. Langendörfer. Detecting Elementary Arm Movements by Tracking Upper Limb Joint Angles with MARG Sensors. *IEEE Journal of Biomedical and Health Informatics*, 20(4):1088–1099, 2016. doi: 10.1109/JBHI.2015.2431472.

[198] M. Miezal, B. Taetz, and G. Bleser. On inertial body tracking in the presence of model calibration errors. *Sensors*, 16(1132), 2016. doi: 10.3390/s16071132.

[199] C. Roldán-Jiménez and A.I. Cuesta-Vargas. Age-related changes analyzing shoulder kinematics by means of inertial sensors. *Clinical Biomechanics*, 37: 70–76, 2016. doi: 10.1016/j.clinbiomech.2016.06.004.

[200] H. Tannous, D. Istrate, A. Benlarbi-Delai, J. Sarrazin, D. Gamet, M.C. Ho Ba Tho, and T.T. Dao. A new multi-sensor fusion scheme to improve the accuracy of knee flexion kinematics for functional rehabilitation movements. *Sensors*, 16(1914), 2016. doi: 10.3390/s16111914.

[201] C.Y. Chiang, K.H. Chen, K.C. Liu, S.J.P. Hsu, and C.T. Chan. Data collection and analysis using wearable sensors for monitoring knee range of motion after total knee arthroplasty. *Sensors*, 17(418), 2017. doi: 10.3390/s17020418.

[202] M. Djurić-Jovičić, N.S. Jovičić, A. Roby-Brami, M.B. Popović, V.S. Kostić, and A.R. Djordjević. Quantification of finger-tapping angle based on wearable sensors. *Sensors*, 17(203), 2017. doi: 10.3390/s17020203.

[203] M.M.B. Morrow, B. Lowndes, E. Fortune, K.R. Kaufman, and M.S. Hallbeck. Validation of inertial measurement units for upper body kinematics. *Journal of Applied Biomechanics*, 33(3):227–232, 2017. doi: 10.1123/jab.2016-0120.

[204] M. Rose, C. Curtze, J. O'Sullivan, M. El-Gohary, D. Crawford, D. Friess, and J.M. Brady. Wearable Inertial Sensors Allow for Quantitative Assessment of Shoulder and Elbow Kinematics in a Cadaveric Knee Arthroscopy Model. *Arthroscopy - Journal of Arthroscopic and Related Surgery*, 33(12):2110–2116, 2017. doi: 10.1016/j.arthro.2017.06.042.

[205] A.F. Ruiz-Olaya, M. Callejas-Cuervo, and C.N. Lara-Herrera. Wearable low-cost inertial sensor-based electrogoniometer for measuring joint range of motion. *DYNA*, 84(201):180, 2017. doi: 10.15446/dyna.v84n201.59054.

[206] T. Sun, H. Li, Q. Liu, L. Duan, M. Li, C. Wang, Q. Liu, W. Li, W. Shang, Z. Wu, and Y. Wang. Inertial Sensor-Based Motion Analysis of Lower Limbs for Rehabilitation Treatments. *Journal of Healthcare Engineering*, 2017:1–11, 2017. doi: 10.1155/2017/1949170.

[207] E. Villeneuve, W. Harwin, W. Holderbaum, B. Janko, and R.S. Sherratt. Reconstruction of angular kinematics from wrist-worn inertial sensor data for smart home healthcare. *IEEE Access*, 5:2351–2363, 2017. doi: 10.1109/ACCESS.2016.2640559.

[208] Z.C. Ong, Y.C. Seet, S.Y. Khoo, and S. Noroozi. Development of an economic wireless human motion analysis device for quantitative assessment of human body joint. *Measurement: Journal of the International Measurement Confederation*, 115:306–315, 2018. doi: 10.1016/j.measurement.2017.10.056.

[209] P. Muller, M.A. Begin, T. Schauer, and T. Seel. Alignment-Free, Self-Calibrating Elbow Angles Measurement Using Inertial Sensors. *IEEE Journal of Biomedical and Health Informatics*, 21(2):312–319, 2017. doi: 10.1109/JBHI.2016.2639537.

[210] E. Allseits, K.J. Kim, C. Bennett, R. Gailey, I. Gaunaurd, and V. Agrawal. A novel method for estimating knee angle using two leg-mounted gyroscopes for continuous monitoring with mobile health devices. *Sensors*, 18(9), 2018. doi: 10.3390/s18092759.

[211] G. Bleser, B. Taetz, M. Miezal, C.A. Christmann, D. Steffen, and K. Regenspurger. Development of an Inertial Motion Capture System for Clinical Application. *i-com*, 16(2):113–129, 2017. doi: 10.1515/icom-2017-0010.

[212] R.A. Fineman, T.M. McGrath, D.G. Kelty-Stephen, A.F.J. Abercromby, and L.A. Stirling. Objective Metrics Quantifying Fit and Performance in Spacesuit Assemblies. *Aerospace Medicine and Human Performance*, 89(11):985–995, 2018. doi: 10.3357/amhp.5123.2018.

[213] T. McGrath, R. Fineman, and L. Stirling. An auto-calibrating knee flexion-extension axis estimator using principal component analysis with inertial sensors. *Sensors*, 18(1882), 2018. doi: 10.3390/s19071504.

[214] Tobias Zimmermann, Bertram Taetz, and Gabriele Bleser. IMU-to-segment assignment and orientation alignment for the lower body using deep learning. *Sensors*, 18(302), 2018. doi: 10.3390/s18010302.

[215] P. Picerno, A. Cereatti, and A. Cappozzo. Joint kinematics estimate using wearable inertial and magnetic sensing modules. *Gait and Posture*, 28(4):588–595, 2008. doi: 10.1016/j.gaitpost.2008.04.003.

[216] M.C. Bisi, R. Stagni, A. Caroselli, and A. Cappello. Anatomical calibration for wearable motion capture systems: Video calibrated anatomical system technique. *Medical Engineering and Physics*, 37(8):813–819, 2015. doi: 10.1016/j.medengphy.2015.05.013.

[217] L. Gastaldi, P. Di Torino, V.A. Politecnico, D. Torino, and S. Pastorelli. Evaluation of the Performances of Two Wearable Systems for Gait Analysis: A Pilot Study. *International Journal of Applied Engineering Research*, 11(16): 8820–8827, 2016.

[218] A. Alizadegan and S. Behzadipour. Shoulder and elbow joint angle estimation for upper limb rehabilitation tasks using low-cost inertial and optical sensors. *Journal of Mechanics in Medicine and Biology*, 17(2):1750031–1–20, 2017.

[219] W. Teufl, M. Miezal, B. Taetz, M. Fröhlich, and G. Bleser. Validity, test-retest reliability and long-term stability of magnetometer free inertial sensor based 3D joint kinematics. *Sensors*, 18(7), 2018. doi: 10.3390/s18071980.

[220] Wolfgang Teufl, Markus Miezal, Bertram Taetz, Michael Fröhlich, and Gabriele Bleser. Validity of inertial sensor based 3d joint kinematics of static and dynamic sport and physiotherapy specific movements. *PloS one*, 14(2):e0213064, 2019.

[221] J.M. Wilken, K.M. Rodriguez, M. Brawner, and B.J. Darter. Reliability and minimal detectible change values for gait kinematics and kinetics in healthy adults. *Gait and posture*, 35(2):301307, 2012.

[222] S.F. Jones, P.C. Twigg, A.J. Scally, and J.G. Buckley. The gait initiation process in unilateral lower-limb amputees when stepping up and stepping down to a new level. *Clinical Biomechanics*, 20(4):405 – 413, 2005.

[223] BikeFit. Bike fitting tools, 2013. `https://www.bikefit.com/documents/GoniometerInstructions.pdf` (accessed: 09.01.2017).

[224] D.F. DuBose, D.C. Herman, D.L. Jones, S.M. Tillman, J.R. Clugston, A. Pass, J.A. Hernandez, T. Vasilopoulos, M. Horodyski, and T.L. Chmielewski. Lower extremity stiffness changes following concussion in collegiate football players. *Medicine and Science in Sports and Exercise*, 49(1):167172, 2017.

[225] A.P. Lapointe, L.A. Nolasco, A. Sosnowski, E. Andrews, D.N. Martini, R.M. Palmieri-Smith, D.H. Gates, and S.P. Broglio. Kinematic differences during a jump cut maneuver between individuals with and without a concussion history. *International Journal of Psychophysiology*, 2017.

[226] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1):3752, 1987.

[227] S.K. Hanneman. Design, analysis and interpretation of method-comparison studies. *AACN advanced critical care*, 19(2):223–234, 2008.

[228] G. Terzakis, P. Culverhouse, G. Bugmann, S. Sharma, and R. Sutton. A recipe on the parameterization of rotation matrices for non-linear optimization using quaternions. Technical Report MIDAS.SMSE.2012.TR.004, Marine and Industrial Dynamic Analysis School of Marine Science and Engineering, Plymouth University, Plymouth, UK, 2012.

[229] D.G. Altman and J.M. Bland. Measurement in medicine: the analysis of method comparison studies. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 32(3):307–317, 1983.

[230] S.-M. Song and K.J. Waldron. *Machines that walk: the adaptive suspension vehicle*. MIT press, 1989.

[231] R.B. McGhee. Some finite state aspect of legged locomotion. *Mathematical Bioscience*, 3:331–351, 1968.

[232] K.E. Adolph, B. Vereijken, and M.A. Denny. Learning to crawl. *Child Development*, 69(5):1299–1312, 1998.

[233] S.K. Patrick, J.A. Noah, and J.F. Yang. Interlimb coordination in human crawling reveals similarities in development and neural control with quadrupeds. *Journal of Neurophysiology*, 101(2):603613, 2009.

[234] T. Kano, C. Owaki, and A. Ishiguro. Reconsidering inter- and intra-limb coordination mechanisms in quadruped locomotion. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vilamoura, Algarve [Portugal], October 2012.

[235] D. Owaki, T. Kano, K. Nagasawa, A. Tero, and A. Ishiguro. Simple robot suggests physical interlimb communication is essential for quadruped walking. *Journal of the Royal Society Interface*, 10(78):20120669, 2013.

[236] D. Owaki and A. Ishiguro. A quadruped robot exhibiting spontaneous gait transitions from walking to trotting to galloping. *Scientific reports*, 7(1):1–10, 2017.

[237] E.J. McElroy, K.L. Hickey, and S.M. Reilly. The correlated evolution of biomechanics, gait and foraging mode in lizards. *Journal of Experimental Biology*, 211(7):10291040, 2008.

[238] S.K. Patrick, J.A. Noah, and J.F. Yang. Developmental constraints of quadrupedal coordination across crawling styles in human infants. *Journal of Neurophysiology*, 107(11):30503061, 2012.

[239] R.L. Freedland and B.I. Bertenthal. Developmental changes in interlimb coordination: Transition to hands-and-knees crawling. *Psychological Science*, 5(1): 2632, 1994.

[240] M.H. Raibert. *Legged robots that balance*. MIT press, 1986.

[241] M. Hildebrand. The quadrupedal gaits of vertebrates. *Bioscience*, 39:766775, 1989.

[242] R.V. Vitali, S.M. Cain, L.V. Ojeda, M.V. Potter, A.M. Zaferiou, S.P. Davidson, M.E. Coyne, C.L. Hancock, A. Mendoza, L.A. Stirling, and N.C. Perkins. Body-worn imu array reveals effects of load on performance in an outdoor obstacle course. *PloS one*, 14(3):e0214008, 2019.

[243] L. Righetti, A. Nylén, K. Rosander, and A.J. Ijspeert. Kinematic and gait similarities between crawling human infants and other quadruped mammals. *Frontiers in Neurology*, 6(17):1–11, 2015.

[244] P.J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

[245] L. Kaufman and P.J. Rousseeuw. *Finding groups in data: an introduction to cluster analysis*. New York: Wiley, 1990.

[246] V. Kumar, J.K. Chhabra, and D. Kumar. Performance evaluation of distance metrics in the clustering algorithms. *Journal of Computational Science*, 13(1): 3852, 2014.

[247] T. Wannier, C. Bastiaanse, G. Colombo, and V. Dietz. Arm to leg coordination in humans during walking, creeping and swimming activities. *Experimental Brain Research*, 141(3):375379, 2001.

[248] J. Cohen. *Statistical power analysis for the behavioral sciences*. Hillsdale, N.J.: L. Erlbaum Associates, 1988.

[249] Task Committee on Undergraduate Science and Engineering Education. Undergraduate science, mathematics and engineering education: Role for the national science foundation and recommendations for action by other sectors to strengthen collegiate education and pursue excellence in the next generation of u.s. leadership in science and technology. Technical Report NSB 86-100, National Science Board (NSB), Washington, D.C.: National Science Board, 1986.

[250] J.W. Moore. The boyer report. *Journal of Chemical Education*, 75(8):935, 1998.

[251] National Academy of Engineering and National Academy of Engineering. *The Engineer of 2020: Visions of Engineering in the New Century*. The National Academies Press, Washington, DC, 2004. doi: 10.17226/10999.

[252] Transforming Undergraduate Education in Engineering. Phase i: Synthesizing and integrating industry perspectives. Workshop report, American Society of Engineering Education, Arlington, VA, 2013.

[253] Transforming Undergraduate Education in Engineering. Phase ii: Insights from tomorrow's engineers. Workshop report, American Society of Engineering Education, Washington DC, 2017.

[254] Transforming Undergraduate Education in Engineering. Phase iii: Voices on womens participation and retention. Workshop report, American Society of Engineering Education, Washington DC, 2017.

[255] Transforming Undergraduate Education in Engineering. Phase iv: Views of faculty and professional societies. Workshop report, American Society of Engineering Education, Washington DC, 2018.

[256] S.M. Lord, E.J. Berger, N.N. Kellam, E.L. Ingram, D.M. Riley, D.T. Rover, N. Salzman, and J.D. Sweeney. Talking about a revolution: Overview of nsf red projects. In *2017 ASEE Annual Conference and Exposition*, Columbus, OH, June 2017.

[257] M. Stains, J. Harshman, M.K. Barker, S.V. Chasteen, R. Cole, S.E. DeChenne-Peters, M.K. Eagan, J.M. Esson, J.K. Knight, F.A. Laski, M. Levis-Fitzgerald, C.J. Lee, S.M. Lo, L.M. McDonnell, T.A. McKay, N. Michelotti, A. Musgrove, M.S. Palmer, K.M. Plank, T.M. Rodela, E.R. Sanders, N.G. Schimpf, P.M. Schulte, M.K. Smith, M. Stetzer, B. Van Valkenburgh, E. Vinson, L.K. Weir, P.J. Wendel, L.B. Wheeler, and A.M. Young. Anatomy of stem teaching in north american universities. *Science*, 359(6383):1468–1470, 2018. ISSN 0036-8075. doi: 10.1126/science.aap8892.

[258] R.R. Hake. Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1):64–74, 1998.

[259] J. Michael. Where's the evidence that active learning works? *Advances in Physiology Education*, 30(4):159–167, 2006.

[260] S. Freeman, S.L. Eddy, M. McDonough, M.K. Smith, N. Okoroafor, H. Jordt, and M.P. Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111(23):8410–8415, 2014.

[261] M. Borrego, J.E. Froyd, and T.S. Hall. Diffusion of engineering education innovations: A survey of awareness and adoption rates in u.s. engineering departments. *Journal of Engineering Education*, 99(3):185–207, 2010. doi: 10.1002/j.2168-9830.2010.tb01056.x.

[262] C. Henderson and M.H. Dancy. Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics - Physics Education Research*, 3(2):020102, 2007.

[263] M.E. Huba and J.E. Freed. *Learner-Centered Assessment on College Campuses: Shifting the Focus from Teaching to Learning*. Boston: Allyn and Bacon, 2000.

[264] J. Handelsman, D. Ebert-May, R. Beichner, P. Bruns, A. Chang, R. DeHaan, J. Gentile, S. Lauffer, J. Stewart, S.M. Tilghman, and W.B. Wood. Scientific teaching. *Science*, 304(5670):521–522, 2004.

[265] K.A. Friedrich, S.L. Sellers, and J.N. Burstyn. 9: Thawing the chilly climate: Inclusive teaching resources for science, technology, engineering, and math. *To Improve the Academy*, 26(1):133–141, 2008.

[266] J.E. Froyd, M. Borrego, S. Cutler, C. Henderson, and M.J. Prince. Estimates of use of research-based instructional strategies in core electrical or computer engineering courses. *IEEE Transactions on Education*, 56(4):393–399, Nov 2013.

[267] M.J. Prince, M. Borrego, S. Cutler, C. Henderson, and J.E. Froyd. Use of research-based instructional strategies in core chemical engineering courses. *Chemical Engineering Education*, 47(1):2737, 2013.

[268] C.J. Finelli, S.R. Daly, and K.M. Richardson. Bridging the research-to-practice gap: Designing an institutional change plan using local evidence. *Journal of Engineering Education*, 103(2):331–361, 2014.

[269] M. Dancy and C. Henderson. Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, 78(10):1056–1063, 2010. doi: 10.1119/1.3446763.

[270] P.C. Boylan-Ashraf, S.A. Freeman, M.C. Shelley, and K. Özgür. Can students flourish in engineering classrooms? *Journal of STEM Education: Innovations and Research*, 18(1):16–24, 2017.

[271] B.S. Bloom. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, 13(6):4–16, 1984.

[272] A.W. Chickering and Z.F. Gamson. Seven principles for good practice in undergraduate education. Technical Report Bulletin 39: 3-7. ED 282 491. 6 pp. MF-01; PC-01, American Association for Higher Education, Washington, D.C., 1987.

[273] C.C. Bonwell and J.A. Eison. Active learning: Creating excitement in the classroom. Technical Report ASHE-ERIC Higher Education Report No. 1, The George Washington University, School of Education and Human Development, Washington, D.C., 1991.

[274] B.S. Bloom. *Taxonomy of Educational Objectives: The Classification of Educational Goals, 1st Ed.* Taxonomy of educational objectives: The classification of educational goals, 1st ed. Longman Group, Harlow, Essex, England, 1956.

[275] L.W. Anderson and D.R. Krathwohl. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Blooms Taxonomy of Educational Objectives*. New York: Addison Wesley Longman, Inc, 2001.

[276] D.R. Krathwohl. A revision of bloom's taxonomy: an overview. *Theory into Practice*, 41(4):212–218, 2002.

[277] W. Bateman. *Open to Question: The Art of Teaching and Learning by Inquiry*. San Francisco: Jossey-Bass, 1990.

[278] V.S. Lee. *Teaching and Learning through Inquiry*. Sterling, Virginia: Stylus Publishing, 2004.

[279] J.E. Mills and D.F. Treagust. Engineering educationis problem-based or project-based learning the answer? *Australasian journal of engineering education*, 3(2):2–16, 2003.

[280] C. Luckmann. Defining experiential education. *Journal of Experiential Education*, 19(1):67, 1996.

[281] M. Gadola and D. Chindamo. Experiential learning in engineering education: The role of student design competitions and a case study. *International Journal of Mechanical Engineering Education*, 0(0):1–20, 2017. doi: 10.1177/0306419017749580.

[282] A. Hajshirmohammadi. Incorporating experiential learning in engineering courses. *IEEE Communications Magazine*, 55(11):166–169, 2017.

[283] R.E. Flori, M.A. Koen, and D.B. Oglesby. Basic engineering software for teaching ("BEST") dynamics. *Journal of Engineering Education*, 85(1):61–68, 1996.

[284] R.E. Flori, D.B. Oglesby, T.A. Philpot, and N.E. Hubing. Incorporating web-based homework problems in engineering dynamics. In *2002 ASEE Annual Conference and Exposition*, Montreal, Canada, June 2002.

[285] N. Fang and Y. Guo. Interactive computer simulation and animation for improving student learning of particle kinetics. *Journal of Computer Assisted Learning*, 32(5):443–455, 2016. doi: 10.1111/jcal.12145.

[286] N. Fang. An analysis of student experiences with concept mapping in a foundational undergraduate engineering course. *International Journal of Engineering Education*, 34(2(A)):294–303, 2018.

[287] B. Self and R. Redfield. New approaches in teaching undergraduate dynamics. In *2001 ASEE Annual Conference and Exposition*, Albuquerque, NM, June 2001.

[288] M.B. Rubin and E. Altus. An alternative method for teaching dynamics. *International Journal of Engineering Education*, 16(5):447456, 2000.

[289] R.A. Streveler, T.A. Litzinger, R.L. Miller, and P.S. Steif. Learning conceptual knowledge in the engineering sciences: Overview and future research directions. *Journal of Engineering Education*, 97(3):279–294, 2008. doi: 10.1002/j.2168-9830.2008.tb00979.x.

[290] M. Kozhevnikov, M.A. Motes, and M. Hegarty. Spatial visualization in physics problem solving. *Cognitive Science*, 31(4):549–579, 2007. doi: 10.1080/15326900701399897.

[291] J. Trindade, C. Fiolhais, and L. Almeida. Science learning in virtual environments: a descriptive study. *British Journal of Educational Technology*, 33(4): 471–488, 2002. doi: 10.1111/1467-8535.00283.

[292] L.G. Kraige, P.R. Gagnon, and Y. Lin. Motion simulation software for personal computer use in undergraduate mechanics courses. In *1985 ASEE Annual Conference and Exposition*, pages 1569–1574, Atlanta, GA, June 1985.

[293] K. Gramoll. Using working model to introduce design to a freshman engineering course. In *1994 ASEE Annual Conference and Exposition*, pages 1–7, Alberta, Canada, June 1994.

[294] J. Iannelli. Mechanics in action: On the development of interactive computer laboratories for engaging engineering mechanics education. In *1994 TBEEC Annual Conference*, Gatlinburg, TN, Nov 1994.

[295] D.A. Smith, R.G. Jacquot, and D.L. Whitman. Lecture enhancing software for engineering dynamics. In *1994 ASEE Annual Conference and Exposition*, Alberta, Canada, June 1994.

[296] D.L. Whitman, R.G. Jacquot, and D.A. Smith. Dynamic dynamics: Projection software for engineering dynamics. In *IEEE Frontiers in Education Conference (FIE)*, pages 1–6, Washington D.C., Nov 1993.

[297] R. Stanley. An efficient way to increase the engineering students fundamental understanding of particle kinematics and kinetics by using interactive web based animation software. *Computers in Education*, 18:2341, 2008.

[298] B.P. Self, J.M. Widmann, M.J. Prince, and J. Georgette. Inquiry-based learning activities in dynamics. In *2013 ASEE Annual Conference and Exposition*, Atlanta, GA, June 2013.

[299] G.C. Adam, B.P. Self, J.M. Widmann, A. Coburn, and B.N. Saoud. How misconceptions might be repaired through inquiry-based activities. In *2015 ASEE Annual Conference and Exposition*, Seattle, WA, June 2015.

[300] A.A. Ferri and B.H. Ferri. Blended learning in a rigid-body dynamics course using on-line lectures and hands-on experiments. In *2016 ASEE Annual Conference and Exposition*, New Orleans, LA, June 2016.

[301] L.D. Feisel and A.J. Rosa. The role of the laboratory in undergraduate engineering education. *Journal of Engineering Education*, 94(1):121–130, 2005.

[302] M.D. Sharma, I.D. Johnston, H. Johnston, K. Varvell, G. Robertson, A. Hopkins, C. Stewart, I. Cooper, and R. Thornton. Use of interactive lecture demonstrations: A ten year study. *Physics Review Special Topics - Physics Education Research*, 6:020119, 2010. doi: 10.1103/PhysRevSTPER.6.020119.

[303] R.R. Hake. Interactive-engagement methods in introductory mechanics courses. *Physics Education Research*, 74:64–74, 1998.

[304] J.R.V. Flora and A.T. Cooper. Incorporating inquiry-based laboratory experiment in undergraduate environmental engineering laboratory. *Journal of Professional Issues in Engineering Education and Practice*, 131(1):19–25, 2005. doi: 10.1061/(ASCE)1052-3928(2005)131:1(19).

[305] J. Watkins and E. Mazur. Retaining students in science, technology, engineering, and mathematics (stem) majors. *Journal of College Science Teaching*, 42 (5):36–41, 2013.

[306] M.J. Graham, J. Frederick, A. Byars-Winston, A.B. Hunter, and J. Handelsman. Increasing persistence of college students in stem. *Science*, 341(6153): 1455–1456, 2013.

[307] A.L. Zeldin and F. Pajares. Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers. *American Educational Research Journal*, 37(1):215–246, 2000.

[308] R.M. Marra, K.A. Rodgers, D. Shen, and B. Bogue. Women engineering students and self-efficacy: A multi-year, multi-institutional study of women engineering student self-efficacy. *Journal of Engineering Education*, 98(1):27–38, 2009.

[309] P.R. Brown and H.M. Matusovich. Career goals, self-efficacy and persistence in engineering students. In *IEEE Frontiers in Education Conference (FIE)*, pages 1–5, Erie, PA, Oct 2016.

[310] M. Vuong, S. Brown-Welty, and S. Tracz. The effects of self-efficacy on academic success of first-generation college sophomore students. *Journal of College Student Development*, 51(1):50–64, 2010.

[311] F. Pajares and M.D. Miller. The role of self-efficacy and self-concept beliefs in mathematical problem-solving: A path analysis. *Journal of Educational Psychology*, 86:193–203, 1994.

[312] C.J. Ballen, C. Wieman, S. Salehi, J.B. Searle, and K.R. Zamudio. Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learning. *CBELife Sciences Education*, 16(4):ar56, 2017.

[313] N. Fang. Correlation between students motivated strategies for learning and academic achievement in an engineering dynamics course. *Global Journal of Engineering Education*, 16(1):6–12, 2014.

[314] F. Pajares and H.D. Schunk. Self-beliefs and school success: Self-efficacy, self-concept, and school achievement. In R. Riding and S. Rayner, editors, *Perception*, pages 239–266. London: Ablex, 2001.

[315] L. Springer, M.E. Stanne, and S.S. Donovan. Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of educational research*, 69(1):21–51, 1999.

[316] D.W. Johnson, R.T. Johnson, and K.A. Smith. Cooperative learning returns to college what evidence is there that it works? *Change: The Magazine of Higher Learning*, 30(4):26–35, 1998.

[317] C.H. Crouch and E. Mazur. Peer instruction: Ten years of experience and results. *American Journal of Physics*, 69(9):970–977, 2001.

[318] L. Deslauriers, E. Schelew, and C. Wieman. Improved learning in a large-enrollment physics class. *Science*, 332(6031):862–864, 2011.

[319] J.K. Knight and W.B. Wood. Teaching more by lecturing less. *Cell Biology Education*, 4(4):298–310, 2005.

[320] P.H. Kvam. The effect of active learning methods on student retention in engineering statistics. *The American Statistician*, 54(2):136–140, 2000.

[321] M.C. Hersam, M. Luna, and G. Light. Implementation of interdisciplinary group learning and peer assessment in a nanotechnology engineering course. *Journal of Engineering Education*, 93(1):49–57, 2004.

[322] S. Bjarne. Teaching engineering dynamics by use of peer instruction supported by an audience response system. *European Journal of Engineering Education*, 36(5):413–423, 2011. doi: 10.1080/03043797.2011.602185.

[323] H. Fencl and K. Scheel. Engaging students. *Journal of College Science Teaching*, 35(1):20–24, 2005.

[324] D.H. Schunk and C.A. Mullen. Self-efficacy as an engaged learner. In *Handbook of Research on Student Engagement.*, pages 219–235. Springer Science + Business Media, New York, NY, US, 2012.

[325] R.W. Lent, S.D. Brown, M.R. Gover, and S.K. Nijjer. Cognitive assessment of the sources of mathematics self-efficacy: A thought-listing analysis. *Journal of Career Assessment*, 4(1):33–46, 1996.

[326] M.A. Hutchison, D.K. Follman, M. Sumpter, and G.M. Bodner. Factors influencing the self-efficacy beliefs of first-year engineering students. *Journal of Engineering Education*, 95(1):39–47, 2006.

[327] D.H. Schunk. Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist*, 25(1):71–86, 1990.

[328] B. Weiner. A theory of motivation for some classroom experiences. *Journal of Educational Psychology*, 71:3–25, 1979.

[329] J.C. Dunlap. Problem-based learning and self-efficacy: How a capstone course prepares students for a profession. *Educational Technology Research and Development*, 53(1):65–83, 2005. doi: 10.1007/BF02504858.

[330] M. Hoit and M. Ohland. Impact of a discipline-based introduction to engineering course on improving retention. *Journal of Engineering Education*, 87(1): 7985, 1998.

[331] M. Lorenzo, C.H. Crouch, and E. Mazur. Reducing the gender gap in the physics classroom. *American Journal of Physics*, 74:118122, 2006.

[332] S.J. Pollock, N.D. Finkelstein, and L.E. Kost. Reducing the gender gap in the physics classroom: How sufficient is interactive engagement? *Physics Review Special Topics Physics Education Research*, 3:010107, 2007.

[333] D.C. Haak, J. Hille Ris Lambers, E. Pitre, and S. Freeman. Increased structure and active learning reduce the achievement gap in introductory biology. *Science*, 332(6034):1213–1216, 2011.

[334] E. Seymour and N.M. Hewitt. *Talking About Leaving: Why Undergraduates Leave the Sciences*. Westview Press, Boulder, CO, 1997.

[335] R.M. Felder, G.N. Felder, and E.J. Dietz. A longitudinal study of engineering student performance and retention. v. comparisons with traditionally-taught students. *Journal of Engineering Education*, 87(4):469–480, 1998. doi: 10.1002/j.2168-9830.1998.tb00381.x.

[336] C.H. Liu and R. Matthews. Vygotsky's philosophy: Constructivism and its criticisms examined. *International Education Journal*, 6(3):386–399, 2005.

[337] D.C. Phillips. The good, the bad, and the ugly: The many faces of constructivism. *Educational researcher*, 24(7):5–12, 1995.

[338] J. Dewey. *Experience and education*. New York: Simon & Schuster, 1938.

[339] E. Von Glasersfeld. Cognition, construction of knowledge, and teaching. *Synthese*, 80:121140, 1989.

[340] T.G. Roberts. An interpretation of dewey's experiential learning theory. Technical Report ED481922, ERIC, 2003.

[341] G.M. Bodner. Constructivism: A theory of knowledge. *Journal of Chemical Education*, 63(10):873–878, 1986. doi: 10.1021/ed063p873.

[342] J. Piaget and D. Elkind. *Six psychological studies*, volume 462. Vintage Books, 1968.

[343] S. Harlow, R. Cummings, and S.M. Aberasturi. Karl popper and jean piaget: A rationale for constructivism. *The Educational Forum*, 71(1):41–48, 2007. doi: 10.1080/00131720608984566.

[344] L.S. Vygotsky. Interaction between learning and development. *Readings on the development of children*, 23(3):34–41, 1978.

[345] L.S. Vygotsky. *Mind in society: The development of higher psychological processes*. Harvard university press, 1980.

[346] P.A. Ertmer and T.J. Newby. Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly*, 6(4):50–72, 1993.

[347] A.E. Rhodes and T.G. Rozell. A constructivist approach to e-text design for use in undergraduate physiology courses. *Advances in Physiology Education*, 39(3):172–180, 2015. doi: 10.1152/advan.00011.2015.

[348] L.B. Bosman, K.L. Chelberg, and S.A. Fernhaber. Introduction to engineering: a constructivist-based approach to encourage engagement and promote accessibility. *Global Journal of Engineering Education*, 19(3):237–242, 2017.

[349] S. Gill and D. Wright. A hypercard based environment for the constructivist teaching of newtonian physics. *British Journal of Educational Technology*, 25 (2):135–146, 1994. doi: 10.1111/j.1467-8535.1994.tb00099.x.

[350] B.D. Coller. A video game for teaching dynamic systems amp; control to mechanical engineering undergraduates. In *2010 American Control Conference*, pages 390–395, June 2010. doi: 10.1109/ACC.2010.5530794.

[351] T.A. Litzinger, L.R. Lattuca, R.G. Hadgraft, and W.C. Newstetter. Engineering education and the development of expertise. *Journal of Engineering Education*, 100(1):123–150, 2011.

[352] B.L. Wiggins, S.L. Eddy, D.Z. Grunspan, and A.J. Crowe. The icap active learning framework predicts the learning gains observed in intensely active classroom experiences. *AERA Open*, 3(2):1–14, 2017. doi: 10.1177/2332858417708567.

[353] X. Wang, D. Yang, M. Wen, K. Koedinger, and C.P. Rosé. Investigating how student's cognitive behavior in mooc discussion forums affect learning gains. In *2015 International Conference on Educational Data Mining*, Madrid, Spain, June 2015.

[354] M. Menekse and M.T.H. Chi. The role of collaborative interactions versus individual construction on students learning of engineering concepts. *European Journal of Engineering Education*, pages 1–24, 2018. doi: 10.1080/03043797.2018.1538324.

[355] W.M. Rodgers, D. Markland, A.-M. Selzler, T.C. Murray, and .M. Wilson. Distinguishing perceived competence and self-efficacy: An example from exercise. *Research Quarterly for Exercise and Sport*, 85(4):527–539, 2014. doi: 10.1080/02701367.2014.961050.

[356] S.L. Hanson. Success in science among young african american women: The role of minority families. *Journal of Family Issues*, 28(1):3–33, 2007.

[357] Z. Hazari, R.H. Tai, and P.M. Sadler. Gender differences in introductory university physics performance: The influence of high school physics preparation and affective factors. *Science Education*, 91(6):847–876, 2007.

[358] C.M. Steele and J. Aronson. Stereotype threat and the intellectual test performance of african americans. *Journal of personality and social psychology*, 69 (5):797–811, 1995.

[359] R.W. Lent, S.D. Brown, and K.C. Larkin. Relation of self-efficacy expectations to academic achievement and persistence. *Journal of Counseling Psychology*, 31(3):356–362, 1984.

[360] E. Bresó, W.B. Schaufeli, and M. Salanova. Can a self-efficacy-based intervention decrease burnout, increase engagement, and enhance performance? a quasi-experimental study. *Higher Education*, 61(4):339–355, Apr 2011. doi: 10.1007/s10734-010-9334-6.

[361] S.G. Brainard and L. Carlin. A six-year longitudinal study of undergraduate women in engineering and science. *Journal of Engineering Education*, 87(4): 369375, 1998.

[362] R.M. Marra, C. Moore, M. Schuurman, and B. Bogue. Assessing women in engineering (awe): Assessment results on women engineering students beliefs. In *2004 ASEE Annual Conference and Exposition*, Salt Lake City, UT, June 2004.

[363] R.V. Vitali, N.C. Perkins, and C.J. Finelli. Introduction and assessment of i-Newton for the engaged learning of engineering dynamics. In *2018 ASEE Annual Conference and Exposition*, Salt Lake City, UT, June 2018.

[364] R.V. Vitali, N.C. Perkins, and C.J. Finelli. Incorporating IMU technology to demonstrate concepts in undergraduate dynamics courses. In *2018 ASEE Annual Conference and Exposition*, Salt Lake City, UT, June 2018.

[365] R.V. Vitali, N.C. Perkins, and C.J. Finelli. Continued assessment of i-Newton for the engaged learning of engineering dynamics. In *2019 ASEE Annual Conference and Exposition*, Tampa, FL, June 2019.

[366] H.G. Espinosa, J. Lee, J. Keogh, J. Grigg, and D.A. James. On the use of inertial sensors in educational engagement activities. *Procedia Engineering*, 112:262 – 266, 2015. 'The Impact of Technology on Sport VI' 7th Asia-Pacific Congress on Sports Technology, APCST2015.

[367] H.G. Espinosa, J. Lee, and D.A. James. The inertial sensor: A base platform for wider adoption in sports science applications. *Journal of Fitness Research*, 4(1):13–20, 2015.

[368] J. Bradshaw and J. Nicholson. The inertial navigation unit: Teaching navigation principles using a custom designed sensor package. In *2008 ASEE Annual Conference and Exposition*, Pittsburgh, PN, June 2008.

[369] A.S. Brill, J.A. Frank, and V. Kapila. Using mounted smartphones as a platform for laboratory education in engineering. In *2016 ASEE Annual Conference and Exposition*, New Orleans, LA, June 2016.

[370] J.A Frank, A.S Brill, and V. Kapila. Mounted smartphones as measurement and control platforms for motor-based laboratory test-beds. *Sensors*, 16(8): 1331, 2016.

[371] C. Chen, C. Pomalaza-Raez, S. Hendershot, M. Oo, and A. Hilton. A wearable wireless system for unobtrusive measurement of human motion. In *2009 ASEE Annual Conference and Exposition*, Austin, TX, June 2009.

[372] S.S. Jordan and M. Lande. Practicing needs-based, human-centered design for electrical engineering project course innovation. In *2012 ASEE Annual Conference and Exposition*, San Antonio, TX, June 2012.

[373] T. Nordenholz. Rigid body dynamics in the mechanical engineering laboratory. In *2008 ASEE Annual Conference and Exposition*, Pittsburgh, PN, June 2008.

[374] S. Shayesteh, M.E. Rizkalla, and M. El-Sharkaway. Curriculum innovations through advancement of mems/nems and wearable devices technologies. In *2017 ASEE Annual Conference and Exposition*, Columbus, OH, June 2017.

[375] S. Bevill and K. Bevill. Smartphone-based measurement of acceleration: Development of a smartphone application for use in an engineering dynamics course. In *2015 ASEE Annual Conference and Exposition*, Seattle, WA, June 2015.

[376] D. Hestenes, M. Wells, and G. Swackhamer. Force concept inventory. *The Physics Teacher*, 30(3):141–158, 1992.

[377] D. O'Connor. Investigations into engineering dynamics theory: A student-led project to utilize smartphone technology. In *2017 ASEE Annual Conference and Exposition*, Columbus, OH, June 2017.

[378] D.F. Treagust. Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10: 159169, 1988.

[379] G.W. Fulmer, H.E. Chu, D.F. Treagust, and K. Neumann. Is it harder to know or to reason? analyzing two-tier science assessment items using the rasch measurement model. *Asia-Pacific Science Education*, 1(1), 2015.

[380] R.J. Osborne and J.K. Gilbert. A technique for exploring the students view of the world. *Physics Education*, 50:376–379, 1980.

[381] M.A. Nelson. Oral assessments: Improving retention, grades, and understanding. *PRIMUS*, 21(1):47–61, 2011.

[382] M. Watts. Exploring pupils alternative frameworks using the interview-about-instances method. In *International Workshop on Problems Concerning Students Representation of Physics and Chemistry Knowledge*, pages 365–386, Ludwigsburg, West Germany, 1981.

[383] Z.D. Kirbulut and O. Geban. Using three-tier diagnostic test to assess students misconceptions of states of matter. *Eurasia Journal of Mathematics, Science and Technology Education*, 10(5):509–521, 2014.

[384] P. Tamir. An alternative approach to the construction of multiple choice test items. *Journal of Biological Education*, 5:305–307, 1971.

[385] N. Jorion, B. Self, K. James, L. Schroeder, L. DiBello, and J. Pellegrino. Classical test theory analysis of the dynamics concept inventory. In *2013 ASEE Annual Conference and Exposition*, pages 1–10, Riverside, CA, June 2013.

[386] B. Finn. Measuring motivation in low-stakes assessments. Technical Report RR-15-19, Educational Testing Service, Princeton, NJ, 2015.

[387] J. Shelley. Using the dci as a continuous process improvement metric for improving student learning outcomes. In *2010 ASEE Annual Conference and Exposition*, Louisville, KT, June 2010.

[388] L.C. Benson, M.K. Orr, S.B. Biggers, W.F. Moss, M.W. Ohland, and S.D. Schiff. Student-centered active, cooperative learning in engineering. *International Journal of Engineering Education*, 26(5):10971110, 2010.

[389] R. Bedillion, R.J. Raisanen, and M. Nizar. Improving transitions between sophomore dynamics and junior dynamic systems courses. In *2014 ASEE Annual Conference and Exposition*, pages 1–10, Indianapolis, IN, June 2014.

[390] Y. Tang and H. Bai. Develop a better way to practice to enhance students' experience in learning dynamics. In *2015 ASEE Annual Conference and Exposition*, Seattle, WA, June 2015.

[391] N. Stites, D.A. Evenhouse, M. Tafur, C.M. Krousgrilland C. Zywicki, A.N. Zissimopoulos, D.B. Nelson, J. DeBoer, J.F. Rhoads, and E.J. Berger. Analyzing an abbreviated dynamics concept inventory and its role as an instrument for assessing emergent learning pedagogies. In *2016 ASEE Annual Conference and Exposition*, New Orleans, LA, June 2016.

[392] S. Ates and E. Catologlu. The effects of students cognitive styles on conceptual understandings and problem-solving skills in introductory mechanics. *Research in Science and Technological Education*, 25(2):167–178, 2007.

[393] K.L. Malone. Correlations among knowledge structures, force concept inventory, and problem-solving behaviors. *Physical Review Special Topics Education Research*, 4(2):020107, 2008.

[394] R.V. Vitali, N.C. Perkins, and C.J. Finelli. Comparing student performance on low-stakes and high-stakes evaluations of conceptual understanding. In *IEEE Frontiers in Education Conference (FIE)*, pages 1–4, Erie, PA, Oct 2018.

[395] M.T. Kane. Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1):1–73, 2013.

[396] J. Vernon, C.J. Finelli, N.C. Perkins, and B. Orr. Piloting i-Newton for the experiential learning of dynamics. In *2015 ASEE Annual Conference and Exposition*, pages 1–9, Seattle, WA, June 2015.

[397] J.C. Nunnally and L. Bernstein. *Psychometric Theory*. McGraw-Hill, New York, 1994.

[398] L.M. Harvill. Standard error of measurement. *Educational Measurement: Issues and Practice*, 10(2):33–41, 1991.

[399] L.A. Clark and D. Watson. Constructing validity: Basic issues in objective scale development. *Psychological assessment*, 7(3):309–319, 1995.

[400] J.W. Pellegrino, L.V. DiBello, and S.P. Brophy. The science and design of assessment in engineering education. *Cambridge Handbook of Engineering Education Research*, pages 571–598, 2014.

[401] J. Neyman and E.S. Pearson. On the use and interpretation of certain test criteria for purposes of statistical inference: Part i. *Biometrika*, 20(1/2):175–240, 1928.

[402] J. Neyman and E.S. Pearson. On the use and interpretation of certain test criteria for purposes of statistical inference: Part ii. *Biometrika*, 20(3/4):263–294, 1928.

[403] J. Neyman and E.S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706):289–337, 1933.

[404] S. Jiang, C. Wang, and D.J. Weiss. Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in Psychology*, 7:109, 2016.

[405] A. Şahin and D. Anil. The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory and Practice*, 17(1):321–335, 2017.

[406] D. Rizopoulos. ltm: An r package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17(5):1–25, 2006. URL `http://www.jstatsoft.org/v17/i05/`.

[407] R. Theobald and S. Freeman. Is it the intervention or the students? using linear regression to control for student characteristics in undergraduate stem education research. *CBELife Sciences Education*, 13(1):41–48, 2014.

[408] J.D. Marx and K. Cummings. Normalized change. *American Journal of Physics*, 75(1):87–91, 2007.

[409] A.J. Neville. Problem-based learning and medical education forty years on. *Medical Principles and Practice*, 18(1):1–9, 2009.

[410] M.A. Zahid, R. Varghese, A.M. Mohammed, and A.K. Ayed. Comparison of the problem based learning-driven with the traditional didactic-lecture-based curricula. *International journal of medical education*, 7:181, 2016.

[411] H.M. Inglis and C. Sandrock. Using the dynamics concept inventory to assess teaching effectiveness. In *2nd Biennial Conf. of South African Society for Engineering Education*, pages 304–306, Cape Town, South Africa., June 2013.

[412] T. Mays, K. Bower, K. Settle, and B. Mitchell. Using concept-oriented example problems to improve student performance in a traditional dynamics course. In *2007 ASEE Annual Conference and Exposition*, Honolulu, HI, June 2007.

[413] Y. Tang. Facilitating learning of projectile problems with a unified approach. In *2017 ASEE Annual Conference and Exposition*, Columbus, OH, June 2017.

[414] E.A. Cech. Ideological wage inequalities? the technical/social dualism and the gender wage gap in engineering. *Social Forces*, 91(4):1147–1182, 2013.

[415] K.N. Smith and J.G. Gayles. girl power: Gendered academic and workplace experiences of college women in engineering. *Social Sciences*, 7(1), 2018. doi: 10.3390/socsci7010011.

[416] L.A. Meadows and D. Sekaquaptewa. The influence of gender stereotypes on role adoption in student teams. In *2013 ASEE Annual Conference and Exposition*, Atlanta, GA, June 2013.

[417] L.J. Hirshfield. Equal but not equitable: Self-reported data obscures gendered differences in project teams. *IEEE Transactions on Education*, 61(4):305–311, 2018.

[418] C. Seron, S.S. Silbey, E. Cech, and B. Rubineau. Persistence is cultural: Professional socialization and the reproduction of sex segregation. *Work and Occupations*, 43(2):178–214, 2016. doi: 10.1177/0730888415618728.

[419] D.A. Cheng. Effects of class size on alternative educational outcomes across disciplines. *Economics of Education Review*, 30:980–990, 2011.

[420] G. McDonald. Does size matter? The impact of studentstaff ratios. *Journal of Higher Education Policy and Management*, 35(6):652–667, 2013.

[421] D. Bradley, P. Noonan, H. Nugent, and B. Scales. Review of australian higher education. Final report, Department of Education, Employment and Workplace Relations, Commonwealth of Australia, Canberra, 2008.

[422] L.S. Toth and L.G. Montagna. Class size and achievement in higher education: A summary of current research. *College Student Journal*, 36(2):253261, 2002.

[423] N.A. Gillespie, M. Walsh, A.H. Winefield, J. Dua, and C. Stough. Occupational stress in universities: Staff perceptions of the causes, consequences and moderators of stress. *Work & Stress*, 15(1):5372, 2001.

[424] R. Schwarzer and M. Jerusalem. *Generalized Self-Efficacy scale*. Windsor, UK: NFER-NELSON, 1995.

[425] U. Scholz, B. Gutiérrez Do na, S. Sud, and R. Schwarzer. Is general self-efficacy a universal construct? psychometric findings from 25 countries. *European Journal of Psychological Assessment*, 18(3):242–251, 2002.

[426] A. Bandura. *Guide for constructing self-efficacy scales*, volume 5. Greenwich, CT: Information Age Publishing, 2006.

[427] N.E. Betz and G. Hackett. The relationship of mathematics self-efficacy expectations to the selection of science-based college majors. *Journal of Vocational behavior*, 23(3):329–345, 1983.

[428] G. Morán-Soto and L. Benson. Relationship of mathematics self-efficacy and competence with behaviors and attitudes of engineering students with poor mathematics. *International Journal of Education in Mathematics, Science and Technology*, 6(3):200–220, 2018.

[429] A.R. Carberry, H.-S. Lee, and M.W. Ohland. Measuring engineering design self-efficacy. *Journal of Engineering Education*, 99(1):71–79, 2010.

[430] P.R. Pintrich, D.A.F. Smith, T. Garcia, and W.J. McKeachie. A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). Technical Report 91-B-004, Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning, 1991.

[431] P.R. Pintrich, D.A.F. Smith, T. Garcia, and W.J. McKeachie. Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measure*, 53(3):801–813, 1993.

[432] M.M. Nauta, J.H. Kahn, J.W. Angell, and E.A. Cantarelli. Identifying the antecedent in the relation between career interests and self-efficacy: Is it one, the other, or both? *Journal of Counseling Psychology*, 49(3):290–301, 2002.

[433] N.E. Betz, L.W. Harmon, and F.H. Borgen. The relationships of self-efficacy for the holland themes to gender, occupational group membership, and vocational interests. *Journal of Counseling Psychology*, 43(1):90–98, 1996.

[434] R.W. Lent, H.-B. Sheu, D. Singley, J.A. Schmidt, L.C. Schmidt, and C.S. Gloster. Longitudinal relations of self-efficacy to outcome expectations, interests, and major choice goals in engineering students. *Journal of Vocational Behavior*, 73(2):328–335, 2008. doi: https://doi.org/10.1016/j.jvb.2008.07.005.

[435] R.W. Lent, M.J. Miller, P.E. Smith, B.A. Watford, K. Hui, and R.H. Lim. Social cognitive model of adjustment to engineering majors: Longitudinal test across gender and race/ethnicity. *Journal of Vocational Behavior*, 86:77–85, 2015. doi: https://doi.org/10.1016/j.jvb.2014.11.004.

[436] J. Payton, T. Barnes, K. Buch, A. Rorrer, and H. Zuo. The effects of integrating service learning into computer science: an inter-institutional longitudinal study. *Computer Science Education*, 25(3):311–324, 2015. doi: 10.1080/08993408.2015.1086536.

[437] I. Ajzen. *From intentions to actions: A theory of planned behavior*. Springer, 1985.

[438] I. Ajzen. The theory of planned behavior. *Organ. Behav. Hum. Decis. Process*, 50(2):179–211, 1991.

[439] L.R.M. Hausmann, J.W. Schofield, and R.L. Woods. Sense of belonging as a predictor of intentions to persist among african american and white first-year college students. *Research in Higher Education*, 48(7):803–839, 2007. doi: 10.1007/s11162-007-9052-9.

[440] D.K. Hatch and C.E. Garcia. Academic advising and the persistence intentions of community college students in their first weeks in college. *The Review of Higher Education*, 40(3):353–390, 2017.

[441] O. Eris, D. Chachra, H.L. Chen, S. Sheppard, L. Ludlow, C. Rosca, T. Bailey, and G. Toye. Outcomes of a longitudinal administration of the persistence in engineering survey. *Journal of Engineering Education*, 99(4):371–395, 2010.

[442] V.E. Wheeless, P.L. Witt, M. Maresh, M.C. Bryand, and P. Schrodt. Instructor credibility as a mediator of instructor communication and students' intent to persist in college. *Communication Education*, 60(3):314–339, 2011. doi: 10.1080/03634523.2011.555917.

[443] T.N. Robinson. Identity as a mediator of institutional integration variables in the prediction of undergraduate persistence intentions. *Journal of Adolescent Research*, 18(1):3–24, 2003. doi: 10.1177/0743558402238274.

[444] W.B. Davidson, H.P. Beck, and M. Milligan. The college persistence questionnaire: Development and validation of an instrument that predicts student attrition. *Journal of College Student Development*, 50(4):373–390, 2009.

[445] J.D. Pugh, J.H. Cramer, S. Slatyer, D.E. Twigg, and M. Robinson. Adaptation and pretesting of the college persistence questionnaire v3 (short form) for measuring intention to persist among aboriginal diploma of nursing students. *Nurse Education Today*, 61:162–168, 2018. doi: https://doi.org/10.1016/j.nedt.2017.11.021.

[446] S. Hidi. Interest and its contribution as a mental resource for learning. *Review of Educational Research*, 60(4):549–571, 1990. doi: 10.3102/00346543060004549.

[447] S. Hidi and K.A. Renninger. The four-phase model of interest development. *Educational Psychologist*, 41(2):111–127, 2006. doi: 10.1207/s15326985ep4102_4.

[448] U. Schiefele, A. Krapp, and A. Winteler. Interest as a predictor of academic achievement: A meta-analysis of research. In *The Role of Interest in Learning and Development*, chapter 8, pages 183–211. Lawrence Erlbaum Associates, Inc, 1992.

[449] J.M. Harackiewicz and C.S. Hulleman. The importance of interest: The role of achievement goals and task values in promoting the development of interest. *Social and Personality Psychology Compass*, 4(1):42–52, 2010.

[450] B.P. Self and J.M. Widmann. Demo or hands-on? a crossover study on the most effective implementation strategy for inquiry-based learning activities. In *2017 ASEE Annual Conference and Exposition*, Columbus, Ohio, June 2017.

[451] J.P. Concannon and L.H. Barrow. A cross-sectional study of engineering students' self-efficacy by gender, ethnicity, year, and transfer status. *Journal of Science Education and Technology*, 18(2):163–172, 2009.

[452] J.P. Concannon and L.H. Barrow. A reanalysis of engineering majors' self-efficacy beliefs. *Journal of Science Education and Technology*, 21(6):742–753, 2012.

[453] J.P. Concannon and L.H. Barrow. Men's and women's intentions to persist in undergraduate engineering degree programs. *Journal of Science Education and Technology*, 19(2):133–145, Apr 2010. ISSN 1573-1839. doi: 10.1007/s10956-009-9187-x.

[454] K.L. Jordan, S. Amato-Henderson, S.A. Sorby, and T.L. Haut Donahue. Are there differences in engineering self-efficacy between minority and majority students across academic levels? In *2011 ASEE Annual Conference and Exposition*, Vancouver, Canada, June 2011.

[455] J. Ranalli and J. Ritzko. Assessing the impact of video game based design projects in a first year engineering design course. In *IEEE Frontiers in Education Conference (FIE)*, pages 530–534, Oklahoma, OK, Oct 2013.

[456] H. Gaikwad and S.S. Kulkarni. Self-efficacy in undergraduate women in engineering - A case study. *Journal of Engineering Education Transformations*, 30 (1):82–86, 2016.

[457] N.A. Mamaril, E.L. Usher, C.R. Li, D.R. Economy, and M.S Kennedy. Measuring undergraduate students engineering self-efficacy: A validation study. *Journal of Engineering Education*, 105:366–395, 2016. doi: 10.1002/jee.20121.

[458] R.B. Kline. *Principles and Practice of Structural Equation Modeling*. New York: The Guilford, 2005.

[459] S.Y. Yoon and S.A. Sorby. Rescaling the longitudinal assessment of engineering self-efficacy v3.0 for undergraduate engineering students. *Journal of Psychoeducational Assessment*, 0(0):1–13, 2019. doi: 10.1177/0734282919830564.

[460] Y. Rosseel. lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2):1–36, 2012. URL `http://www.jstatsoft.org/v48/i02/`.

[461] T.A. Brown. *Confirmatory factor analysis for applied research*. Guilford Publications, 2015.

[462] A. Field, J. Miles, and Z. Field. *Discovering statistics using R*. Sage publications, 2012.

[463] M.J. Eaton and M.H. Dembo. Differences in the motivational beliefs of asian american and non-asian students. *Journal of Educational Psychology*, 89(3): 433, 1997.

[464] C.S. Dweck. Motivational processes affecting learning. *American psychologist*, 41(10):1040, 1986.

[465] R.L. Navarro, L.Y. Flores, H.-S. Lee, and R. Gonzalez. Testing a longitudinal social cognitive model of intended persistence with engineering students across gender and race/ethnicity. *Journal of Vocational Behavior*, 85(1):146–155, 2014.

[466] E.B. Coleman, A.L. Brown, and I.D. Rivkin. The effect of instructional explanations on learning from scientific texts. *The Journal of the Learning Sciences*, 6(4):347–365, 1997.

[467] J.D. Gobert and J.J. Clement. Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 36(1):39–53, 1999.

[468] S. Zhao. Time derivative of rotation matrices: A tutorial. *CoRR*, abs/1609.06088, 2016.