The Mathematics Enthusiast

Manuscript 1574

A Course-Based Undergraduate Research Experience (CURE) in Soccer Analytics

Russell E. Goodman

Katelin Valster

A Course-Based Undergraduate Research Experience (CURE) in Soccer Analytics

Russell E. Goodman and Katelin Valster
Central College, Pella, IA

ABSTRACT: The authors designed an honors seminar course entitled *Sports Analytics: Soccer* to provide students with a course-based undergraduate research experience (CURE) using cutting-edge GPS/accelerometer athlete-tracking devices. This paper offers details of the design and structure of the course, where a goal was for students to contribute to a longitudinal research project of performing a comprehensive performance and load analysis on Division III women's soccer athletes.

Keywords: undergraduate research, sports analytics, soccer, pedagogy

The Mathematics Enthusiast, **ISSN 1551-3440, vol. 19, no. 3**, pp. 751-758

Introduction

Undergraduate research is supported by the Mathematical Association of America and encouraged by many undergraduate institutions of higher learning. Opportunities for students to conduct research outside of their regular classroom experiences, such as during summer programs, are plentiful. However, another way of providing students with research opportunities is through course-based undergraduate research experiences (CUREs). The authors designed a one-credit, fall 2020 course entitled *Sports Analytics: Soccer* to provide students, not only mathematics majors, with a CURE in exploring the analytics of the game of soccer.

> Catalog Course Description: *This course will be an investigation of the role data analysis plays in the sports world, specifically in soccer. Such analysis is referred to as* sports analytics. *Students will be at the forefront of a research project where they will develop, collect and analyze real data coming from the Central College women's soccer team. The focus will be on learning the skills and habits necessary to create a comprehensive performance analysis of those student-athletes. This analysis will help us begin the long-term goal of understanding the physical, mental, and emotional loads placed on Central College student-athletes.*

In designing the course, the instructors articulated a broad "umbrella" project of a comprehensive performance and load analysis of women's soccer athletes at a Division III school such as Central College. Performance and load analysis can encapsulate any number of variables relevant to sport training as an attempt to describe the intensity and volume of a given training session or competition. Further, the reader will note that their course structure aligned strongly with many aspects of CUREs, as defined in [Auch].

While player load data has been considered in other soccer athlete participant groups (see [HBP], [HDCR], [SLK]) there is relative paucity of data at the Division III level. Improved measurements through GPS and accelerometer technology offer new insight to quantify training load, understand responses to training and competition, and improve training prescriptions for both the team and individual athlete. This is particularly true for Division III women's soccer which has the most teams and participants of any other NCAA sponsored sport ([NCAA]). Yet, no known player load quantification or mechanical demands of this sub-group are available in the literature.

Player load can be described as the amount of work performed in a training or competition time period. This can include variables of external load (i.e., distance covered, number of accelerations) or internal load (i.e., heart rate, rating of perceived exertion) and represent examples of mechanical demands during human movement (see [Murr], [TBK]). For example, Catapult Sports uses a method of "summing the squares of each reading and dividing the square root of this value by 100" ([TBK]) in order to quantify their Player Load™ variable.

Recent introduction of GPS and accelerometer technology to team sport settings presents the ability to monitor these mechanical demands, especially external player load, at a more precise level during actual sport play. The differences in sport season length, rules, and resources are important distinctions as they offer varied demands compared to other elite soccer athletes (e.g., high-level club soccer, Division I, or professional play). The research pursued in this course had the potential to positively influence training prescriptions for Division III women's soccer athletes, thereby adding to the growing body of literature regarding player load and data analytics in team sport. In this paper, the reader will learn how the course structure was well-aligned with CURE features such as the use of reputable scientific practices, discovery, and collaboration.

1 About Central College and (Funding for) Devices

Central College is a private, four-year liberal arts college located in Pella, Iowa. Its enrollment is approximately 1,100 students with no graduate-level programs. Further, Central participates in NCAA athletics in the American Rivers Conference at the Division III level. Institutions competing at the Division III level commit to not offering athletic scholarships to any student-athletes.

A key aspect of this course and research experience was the acquisition of the appropriate technology for the expected data collection. The instructors planned for a class of 10-15 students who would self-identify into teams of 2-4 on projects of their own choosing and that were deemed feasible with the available technology. As a result, the instructors focused on acquiring a collection of Catapult PLAYERTEK

devices that measure GPS location, physical impacts, sprint distance, and other related accelerometer measurements. Access and utilization of the Catapult devices allowed the students to think broadly and with autonomy about an aspect of physical load that appealed most to them for their research project.

The instructors applied for, and received, funding from the Moore Family Foundation Faculty Development Program for Teaching in an amount that allowed for the purchase of 20 such devices along with the necessary accessories such as vests, chargers, carrying case, etc.

2 Description of Intended Students' Research Experience

The intention was to have students engage in the entire research process, from a review of the literature, development of the research question, method design, data collection, data analysis, and presentation of the information in a professional setting. The course was intentionally designed for "honors-eligible" students in order to most easily facilitate the intense and fast-paced research experience. At Central College, first year students are honors-eligible if they had a high school GPA of at least 3.75 with a minimum score of 28 on the ACT, or if they were in the top 5% of their high school graduating class. Returning students are honors-eligible if their cumulative GPA is at least 3.5. As a result, the course attracted undeclared first-year students along with students from a variety of majors (e.g., exercise science), not exclusively mathematics. The course fulfilled progress within the honors program at the institution, but did not count toward core or major credit distinctions.

The students were encouraged to collaborate, consider minimal invasiveness of their project on the athletic season, and problem-solve during the research process, especially as sports were under daily modifications due to the COVID-19 pandemic. An additional goal of the experience was to introduce students to the Institutional Review Board (IRB) application process, as that is a common and important step in research involving human subjects.

One can see from the course timeline and ensuing description (Section 4 and Appendix, below) that the course aligned with the definition of a CURE in a variety of ways. As a result, the students had a developmental and transformative experience.

- Use of scientific practices. Class time was devoted to a "what is a good research question?" discussion which quite naturally led to the students' work of designing their studies. As the process evolved, students were responsible for the logistics of data collection, data analysis, satisfying Central College's IRB process, and most importantly, embracing the messiness of the research process and the data it provided. Even under ideal conditions, a study rarely goes smoothly. The COVID-19 pandemic presented unprecedented challenges to both the student researchers and the college soccer team. Learning to be flexible and adaptable regarding team training times, weather conditions, researcher availability, etc. was essential.

- Discovery and relevant work. A key aspect of a CURE is for the outcome of an investigation to be unknown to the researcher and in this course that was certainly the case. The students in the course came from a variety of backgrounds with not all even knowing the structure and rules of the game of soccer. Further, the instructors emphasized the opportunity to contribute to a gap in the literature regarding physiology of soccer and the students embraced this opportunity. While their work was unlikely to end up in a publication (inappropriate or insufficient data), the students engaged in this opportunity in good faith and with vigor.

- Collaboration. Good science often involves teams of researchers who bring an array of skills and mindsets to bear on a study. In this course, student research projects were individualized, but the instructors worked to find commonality among the projects to allow for productive peer discussion and critique during the course. Class periods were occasionally structured to allow the students to review one another's research questions, research proposals, IRB proposals, and research progress reports. Further, the instructors emphasized the shared responsibilities of all researchers due to the data collection process, working to ensure the data collection process was executed with integrity and accuracy.

3 Examples of Student Research Work

This section is designed to share examples of student work coming from three specific assignments: a podcast review early in the semester, a research poster, and a final project report.

Early in the course, students were introduced to notions of soccer, exercise physiology, and sports analytics. The instructors asked the students to carefully listen to a soccer panel discussion, hosted by Grant Wahl, from the 2020 MIT Sloan Sports Analytics Conference [Wahl]. Students then wrote a reflection, considering what they found to be the most interesting ideas coming from the panel, what notions were new to them, and what ideas might impact the work in the class. This assignment was positioned in parallel with students working through literature reviews and development of a research question. One student's response to this podcast focused on aspects of the panel discussion related to GPS tracking technology. He made a nuanced connection between the podcast conversation and his topic of "team dominance" in the thirds of the field: defending, middle, and attacking. This assignment set the tone, in a broad way, for getting the students up to speed on fundamental, contemporary issues in (soccer) analytics.

As the students pursued their projects, they were expected to create a professional research poster in order to share their work at the (virtual) 2020 Midwest Sports Analytics Meeting (MSAM). A noteworthy example of student work was *Reload and Hang Loose: Analyzing Mentality and Work Rate in Women's Division III Soccer* [Huin]. The researcher's project focused on the connection between "negative events" in a soccer match and the physical evidence that might represent a "bounce-back" effect in order to create a positive event. The student, and the instructors, believe there is promise for this project to develop further in a post-COVID-19 atmosphere when more contact, communication, and interaction with the participants is possible.

Along with producing a professional research poster, the students were expected to write a formal research report. Ideally, the report would have the potential for publication but that was not a driving influence on the assignment. Two examples of such work came from student researchers whose projects were entitled *A Study of Weight Training in Soccer Players* and *The Effect of Impacts on NCAA Women's Soccer Goalkeepers* (see [Albe], [Spoe]) . Through their reports, the researchers shared a common structure of reporting on the research question, methods, preliminary results, reflections, and potential future research on the topic.

4 Timeline of Course

Central College's semesters are 15 weeks in length and *Sports Analytics: Soccer* met once per week, on Mondays. A typical class period involved recapping the previous week's activities and discussions (a week between class periods is a long time!), activities to move the research projects along, and an assignment of tasks for the following week. Courses at Central College are typically three or four credits, typically meeting three times per week. Since this was a one-credit course, our expectations were that the students would put in a minimum of two hours of work outside of class to go along with the one hour of time in class. While further elaboration on the organization of the course is in the next section, this paper's Appendix contains a table outlining the week-to-week activities of the instructors' fall 2020 class.

5 Additional Detail on Course Materials, Activities, and Assignments

The instructors provided articles from the literature to introduce team sport, including soccer, GPS and accelerometer data analytics. Students were then assigned to search and gather additional scholarly articles to support their specific research questions. Articles were shared among the group in order to build a more robust body of literature. Most weekly class sessions centered on discussion from these articles rather than utilizing a formal text due to the quick speed at which the course moved. The Blackboard learning management system was utilized extensively as a common location for all course materials and assignments.

The two instructors have been trained in distinct disciplines but shared common interests and experiences with the sport of soccer. Consequently, the Mathematics professor led the data components of the

course: the data management, introduction of appropriate analysis software, performing data analysis, and data visualization. Analysis software considered were PowerBI, Tableau, R, and Google Sheets. Ultimately, the students used Google Sheets primarily due to its availability, ease of use, and for technical reasons connected to the students' individual computers. The Exercise Science professor offered insight into primary considerations for sport research methods including expert knowledge in exercise physiology and implications for designing feasible research projects.

The mathematical, data-oriented, and statistical content of the course was designed appropriately, given that the students came from a variety of disciplines. The students were introduced to performing an exploratory data analysis (EDA) in order to summarize datasets, discover patterns, spot anomalies, frame their hypotheses, and to check or challenge their assumptions. A mini-lecture offered basic statistical concepts such as using correlation to search for relationships between variables, using visuals like scatterplots and histograms as part of EDA, and fundamentals of hypothesis testing. Additionally, the students needed guidance in navigating the Catapult interface in order to understand, acquire, and prepare the data that the accelerometers were providing. Student work in data preparation included cleaning up player identifications, dates, and subsetting the data as appropriate for their specific research project.

A significant component of the course design was to articulate how much time research design takes before one even gets to the point of data collection. For example, data collection did not start until week eight and initial visualization of data collection through the Catapult interface and data analysis occurred the following week. Yet, students still needed to modify their projects to some degree due to participant responsiveness, team training logistics, and other various reasons. This meant discussion and activities included continuous prompts to focus the research question(s) and problem-solving to collect reputable data. One student's project, for example, relied on intrasquad scrimmages for data collection and analysis, however the practice design for the season (due to COVID-19 restrictions) resulted in short ($<$ 12 minutes) small-sided games (e.g., 7 vs. 7). While not ideal, it allowed for productive conversation and collaboration to adjust to the basic relationship of what the student(s) wanted to study.

Assignments aligned with the typical development of a research project. Discussions began with an introduction to team sport data analytics along with required written responses to an assigned podcast and additional readings. Early in the course, students were expected to start their literature search on topics within their interest area. For example, a student who was interested in the association between strength levels and practice session speed was expected to quickly begin to accumulate reputable information regarding those variables. A student who wanted to understand the biomechanics and physical challenges of goalkeeping needed to search for reputable articles related to those topics. Students were encouraged to use the references in the initially-provided articles as a launching point for their literature review.

Students were asked to develop two or three research questions on which the instructors provided detailed feedback and guidance on the feasibility of their potential research directions. Upon approval of their official research question, students embarked on writing a research proposal which was to include an introduction to their specific topic in soccer data analytics, their research question, and intended methods. Additionally, students were informed of the IRB process and began to write an IRB proposal and informed consent form. While the instructors ultimately submitted an "umbrella" IRB proposal on behalf of the entire class, students were engaged in the process to develop the necessary skills to conduct human research. Both the research proposal and IRB forms were evaluated by the professors via assignment, 1-on-1 meeting, and peer review. The development of the research proposal lasted approximately four weeks; a significant part of the academic semester. Then, students were challenged to think outside their own project as to how collected data could serve all project designs in the course, especially given the circumstance of having fewer accelerometer devices than athletes on the soccer team. For instance, one group was interested in player load of goalkeepers while another wanted to collect whole-team scrimmage data. The instructors challenged the students to consider sampling procedures to serve the needs of all involved projects. Moreover, the instructors encouraged students to contemplate the "choreography" of data collection among projects, including video needs, collection and return of the GPS/accelerometer equipment, and data file sharing.

Data analysis and interpretation was expected of each individual, though time was built in class to brainstorm and collaborate with peers and faculty. After suitable insight was articulated, students began to draft a professional research poster. The instructors spent class time presenting both good and avoidable qualities of research posters to prepare the students for their opportunity to present their work.

Evaluation was provided by faculty through a formal assignment and in-class through peer review. In a final celebration of their work, students were able to present to academics and professionals at MSAM. This opportunity solidified the entire research process from start to finish. The final two weeks of the semester (week 13 and 15) encouraged students to verbally reflect on the research process itself and the course in general and to generate their final research report.

6 Conclusion

This course-based undergraduate research experience, or CURE, was an opportunity for Central College students to conduct research via one of their regular classroom experiences. The authors designed a one-credit, fall 2020 course entitled *Sports Analytics: Soccer* to provide students, not only mathematics majors, with a CURE in exploring the analytics of the game of soccer. The instructors' longitudinal "umbrella" project of a comprehensive performance and load analysis of women's soccer athletes at the Division III level was initiated with this course and the course structure aligned strongly with many aspects of CUREs, as defined in [Auch].

The use of scientific practice in development of an undergraduate research project occurred throughout the semester, a positive outcome, but due to pandemic-related training limitations within the athletic department, the data collected lacked compelling insight into the physical demands on Division III women's soccer players. A few student evaluations noted they appreciated the acquisition of the fundamental research skills. In fact, two students proceeded to earn summer research grants while another two are continuing to have research experiences through senior honors theses.

The course design was intentionally open-ended. That is, student interest drove their research agenda rather than the faculty dictating what the project should or should not be. This challenged the students' creativity and discovery in the research process. However, verbal reflections in the final two class periods helped the faculty members understand they needed to spend more time at the start of the course describing the Catapult technology and its capabilities in order for the students to better discern their research focus.

Additional outcomes for the instructors include further understanding the time it takes for (first-year) students to feel comfortable giving constructive and tactful feedback to their peers, gaining valuable experience driving a collaborative research environment, and anecdotally observing the students' growth in curiosity, formal writing, and maturity.

The instructors wished for the research pursued in this course (and in the future) to have the potential to positively influence training prescriptions for Division III women's soccer athletes, and thereby add to the growing body of literature regarding player load and data analytics in team sport. Equally important is that the use of reputable scientific practices, discovery, and collaboration led to a developmental and transformative experience for the students.

7 Appendix – Weekly Schedule

Week	Class Activity/Content	Assignments
1	Introductions; course overview; the game of soccer: rules, history, teams, tactics; discussion of COVID-19 and our needing to be "flexible and adaptable"; "unknowns versus the 'ideal"'; discuss what makes a good research question	Review Catapult website; one-page reflection on a podcast episode from 2020 MIT SSAC; devise at least 2 research questions; mini-annotated bib with 4+ sources due in two weeks (Labor Day), connected to research questions
2	**no class - Labor Day**	
3	Update on soccer season; discuss research questions and sources investigated; discuss Catapult devices and player load to develop research question ideas; overview of what a research project involves; intro to IRB; look at sample Catapult dataset	Complete PowerBI tutorial; develop project proposal draft; read IRB forms
4	Visit from Associate Dean re: IRB purpose and procedures; discuss exploratory data analysis	Write project IRB proposal; continue collecting project sources; 2 articles and prompts to respond to: wearable technology and player load
5	Peer evaluation and critique of project proposals; Intro to soccer physiology and player load	Project proposal revisions, writing an author's note; schedule 1-on-1 meeting with a professor; professors meet with soccer team to introduce the research, the devices, and gather interested players' names
6	Update on soccer season and professors' meeting with team; work with sample data set (review, clean, and analyze); discuss sampling options	Add content to "shared responsibilities" Google Doc with player responsibilities with devices, student researcher responsibilities and needs in collecting data during training sessions; review and edit IRB proposal
7	Play with data collected to date; continued work on "shared responsibilities" document and discussion of logistics for collecting practice data	Review Catapult technology capabilities; continue the work in cleaning, organizing, and analyzing data collected to date - make progress on projects!
8	**no class - Fall Break**	
9	Review video from practice; adjust protocol and variables (as needed); discuss poster presentations "do's and don'ts"	Students ensure questionnaire data available to all class participants; register for MSAM conference
10	Discuss how to manage and edit data file(s) via Google Drive; identify patterns in collected data; continue to discuss poster expectations; work time	Re-affirm variables needed to appropriately analyze the relationships for analysis; continued work on poster - first draft due; register for MSAM
11	Review and peer-review of posters	Final revisions of posters; prepare for presentation at MSAM
12	Practice poster presentation for MSAM; work day	Attend MSAM and at least one session/talk; write a reflection on presenting and on the session/talk you attended
13	Discussion of MSAM conference; discuss final assignment: research report - due at final exam period	Work on research report - due at final exam period
14	**no class - work day**	
15	**Final Exam period:** final research reports due; semester review and Charty Party (game)	

References

[Albe] Alberti, A. (2021, September 2). *A Study of Weight Training in Soccer Players* [Unpublished Report]. Retrieved from https://tinyurl.com/r6ma9j4m.

[Auch] Auchincloss, L., et. al. (2014). Assessment of Course-Based Undergraduate Research Experiences: A Meeting Report. CBE-Life Sciences Education, 13, pp. 29–40. https://www.lifescied.org/doi/full/10.1187/cbe.14-01-0004

[HBP] Harley, J. A., Barnes, C. A., Portas, M., Lovell, R., Barrett, S., Paul, D., & Weston, M. (2010). Motion analysis of match-play in elite U12 to U16 age-group soccer players. *Journal of Sports Sciences*, 28(13), 1391–1397.

[HDCR] Hill-Haas, S. V., Dawson, B. T., Coutts, A. J., & Rowsell, G. J. (2009). Physiological responses and time-motion characteristics of various small-sided soccer games in youth players. *Journal of Sport Sciences*, 27(1), 1–8.

[Huin] Huinker, L. (2021, September 2). *Reload and Hang Loose: Analyzing Mentality and Workrate in Women's Division III Soccer* [Powerpoint]. Retrieved from https://tinyurl.com/3p93m9ut.

[Murr] Murray, A. (2019). Monitoring training load in American football. *NSCA Coach*, 5(4), 30–35.

[NCAA] NCAA Research (2020, September 1). NCAA Sports Sponsorship and Participation Rates Report. https://ncaaorg.s3.amazonaws.com/research/sportpart/2019-20RES_SportsSponsorshipParticipationRatesReport.pdf

[SLK] Scott, B. R., Lockie, R. G., Knight, T. J., Clark, A. C., & Janse de Jonge, X. A. K. (2013). A comparison of methods to quantify the in-season training load of professional soccer players. *International Journal of Sports Physiology and Performance, 8*, 195–202.

[Spoe] Spoehr, T.X. (2021, September 2). *The Effect of Impacts on NCAA Women's Soccer Goalkeepers* [Unpublished Report]. Retrieved from https://tinyurl.com/nz6h3bah.

[TBK] Theodoropolous, J. S., Bettle, J., & Kosy, J. D. (2020). The use of GPS and inertial devices for player monitoring in team sports: A review of current and future applications. *Orthopedic Reviews, 12:7863*, doi: 10.4081/or.2020.7863.

[Wahl] Wahl, G. (Host). (2020, May 23). *Soccer Analytics: The Beautiful Game Meets the Analytics Edge* [Audio podcast]. Retrieved from https://ssac.podbean.com/e/soccer-analytics-the-beautiful-game-meets-the-analytics-edge/

Department of Mathematics, Central College, Pella, IA 50219, USA
Email address: goodmanr@central.edu

Department of Exercise Science, Central College, Pella, IA 50219, USA
Email address: valsterk@central.edu