A Data-Driven Framework for Skill Representation

Mariah Schrum, Allison Morgan, Deepak Gopinath, Jean Costa, Emily Sumner, Guy Rosman, Tiffany Chen
Human Interactive Driving
Toyota Research Institute
Los Altos, USA
first.last@tri.global

***Abstract*—Understanding and modeling human skill is critical for advancing human-robot and human-AI interaction, particularly in domains requiring nuanced cooperation and long-term personalization. Effective collaboration depends on aligning AI behavior with human capabilities, but quantifying skill is challenging, requiring both task knowledge and expert intuition. This work proposes a data-driven approach to modeling human skill through repeated interactions. Using high-performance driving education (HPDE) as a case study, we synthesize literature and expert insights to identify supervision signals for learning a robust skill representation. Leveraging a dataset of novice and expert drivers, we demonstrate the feasibility of automatically extracting skill representations from track-driving trajectory data. This foundation enables applications such as personalized coaching, skill-targeted challenges, and adaptive robotic interventions. Ultimately, we aim to democratize HPDE by making high-quality instruction more accessible through AI-driven personalization.**

***Index Terms*—skill modeling, long-term personalization, HPDE**

I. Introduction

For an AI system to effectively work with and aid a human during long-term interactions on a complex task, it must estimate the skill of its human collaborator. Without understanding a human's skill, the AI system may inappropriately intervene or make inaccurate assessments, limiting its ability to personalize interactions and adapt to the human's needs.

Often, *skill* and *performance* are used interchangeably [1], but in this work, we make a clear distinction between these two constructs. *Skill* is the underlying capability that enables consistent, adaptable task execution and can develop over time through practice. *Performance*, in contrast, reflects task outcomes (e.g., lap times) influenced by skill and external factors like fatigue or environment familiarity [1]. Accurately modeling skill, separately from performance, is crucial for AI systems to assist, teach, and collaborate effectively [2]. However, skill is a complex, evolving construct requiring expert knowledge and deep task understanding to represent effectively.

We propose a machine-learning-based approach to model both skill and performance in a disentangled way, leveraging both data and domain expertise to derive a representation of human skill that closely aligns with expert practitioners' understanding of the construct. Using high-performance driving education (HPDE) as a case study, our methodology synthesizes findings from sports and motor learning literature as well as supporting evidence gleaned from qualitative interviews with expert coaches to disentangle skill from performance and effectively represent skill in an AI system.

Expert coaches often describe skill in terms of *subskills*—distinct but interrelated components that contribute to overall task proficiency (Section III-A). In HPDE, these may include technical abilities such as car control and race line execution, cognitive aspects like situational awareness and decision-making, and perceptual-motor skills such as gaze ability and reaction time. In our work, we aim to model these subskills explicitly to allow for a finer-grained and interpretable skill representation.

To accurately assess their students' skill levels, expert HPDE coaches observe students during task execution and infer skill using a mental model refined through years of experience. Given observations of a student's driving behavior, coaches use this internal model to estimate the student's skill level, distinguishing between underlying ability and momentary performance fluctuations. Our objective is to enable AI systems to replicate this inference process by learning a mapping of observations to skill representations that can be leveraged to improve downstream interactions such as coaching or assistance. While human coaches develop their mental models by accumulating and interpreting data over years of experience, we aim to collect similar data at scale to train an AI model capable of learning a comparable skill assessment process. Towards this goal, we identify five key steps for modeling skill.

1) Understand subskills in HPDE via expert interviews to inform how our AI system structures skill representation.
2) Identify key *skill metrics* for measuring subskills, providing necessary training signals for the AI system.
3) Collect a dataset of student task execution and metrics to train the AI system to extract skill representations.
4) Verify that the dataset contains sufficient signals to distinguish between skill levels.
5) Leverage the dataset and domain expertise towards data-driven modeling of skill.

In this work, we present results from steps 1-4 and outline plans for step 5 to use the dataset and domain expertise to train a skill representation.

Through this methodology, we pave the way for personalized AI coaching, adaptive robotic assistance, and targeted interventions for skill deficits. Beyond racing, our approach supports skill development in sports and motor domains through data-driven insights. In HPDE, our goal is to bridge

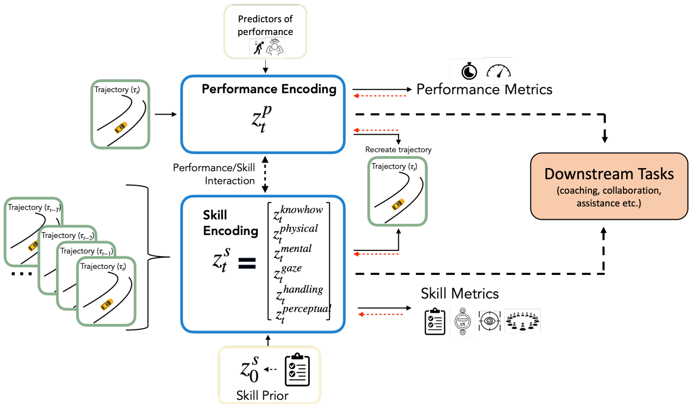

Fig. 1: This figure shows our proposed model architecture. z^s captures skill and z^p captures performance

human expertise and AI, democratizing access to high-quality coaching for more inclusive skill development.

II. Related Work

Prior research on human skill modeling has explored expert heuristics, cognitive models, and data-driven techniques. Knowledge tracing has examined various features for evaluating student performance but lacks the necessary components for motor skill characterization [3]. Traditional driving education relies on subjective instructor evaluations rather than quantitative measures [4], while machine learning approaches leverage behavioral data—such as gaze patterns, control inputs, and trajectory consistency—to represent human ability [5], [6]. Other techniques include co-training performance estimates with teacher imitation [7] and modeling a student's zone of proximal development [8].

Skill assessment techniques extend beyond driving, with applications in medicine for evaluating surgeons [9] and in tutoring systems [10]. Sports science differentiates skill from performance by accounting for external factors like fatigue and stress [1], [11]. Similarly, human-robot interaction (HRI) studies emphasize the need for AI systems to adapt to human skill levels over time, particularly in long-term collaborations [12]. Building on these foundations, our work synthesizes insights from expert interviews and cognitive science to develop a robust, data-driven skill representation.

III. Methodology

Our methodology follows a systematic approach to developing a data-driven representation of human skill, grounded in domain expertise and literature. Below, we outline the key steps involved in learning a representation of skill in HPDE.

A. Expert Insights and Literature Review

To enable AI systems to effectively understand skill, our first objective is to explore how experienced HPDE coaches evaluate skill. We conducted one-hour semi-structured interviews with three seasoned HPDE coaches. Our goals were two-fold: 1) to identify the core subskills that constitute skill in HPDE, and 2) to understand the variables, metrics, and tasks that coaches use to assess these subskills.

During the interviews, we asked coaches questions to gain insights into key subskills in HPDE, how they form initial impressions of a student's skill before entering the vehicle, the metrics they use to assess skill during execution, and what supplementary tasks provide additional evidence of skill. We also explored how coaches distinguish skill from performance in their evaluations. Further details provided in Appendix E.

These insights were further supplemented by a review of cognitive science and sports literature, which provided a theoretical foundation for understanding skill as a multifaceted construct. These interviews and literature review informed the design of our data collection and modeling efforts to ensure that we captured the complexity of skill in a way that aligns with both expert knowledge and empirical data. Details of this literature review are provided in Appendix H.

B. Subskills in HPDE

Based on our interviews with HPDE coaches, we identified six key subskills in HPDE through an inductive coding process. We first extracted all skill-related concepts mentioned by coaches, and then grouped them into recurring themes to form our final subskill categories.

- Know-How: Practical understanding of fundamental principles and concepts, including both declarative knowledge and procedural knowledge.
- Physical: Strength and physical endurance.
- Mental: Ability to handle cognitive load, situational awareness, and emotional regulation.
- Gaze Policy: Visual attention and fixation patterns
- Vehicle Handling: The ability to apply smooth, precise, and intentional control over the vehicle
- Perception: Ability to efficiently process and interpret visual and spatial information and predict outcomes

Consistency, while not a subskill in and of itself, is a fundamental requirement across all subskills. Skilled drivers exhibit consistency by reliably executing subskills over time, minimizing variability, and ensuring repeatable performance.

C. Skill Metrics

Drawing from expert insights and literature, we developed a set of metrics to separate skill levels (i.e., novice and expert). These enable our AI system to assess each subskill and forms the foundation for training our AI system to learn a skill representation. We note this set is not exhaustive — we focus on the metrics we have analyzed thus far, and leave additional factors that further refine skill assessment to future work. We provide a more comprehensive list of metrics in Appendix D. Because performance can be evaluated using simple metrics like lap time, our primary objective is to develop a robust representation of skill—capturing underlying ability, consistency, and cognitive factors that influence learning and improvement.

- Know-How: performance on knowledge test [13]
- Physical: Grip strength [14] and performance on hand-eye and motor skill tests [15] assessed using PsyToolkit.

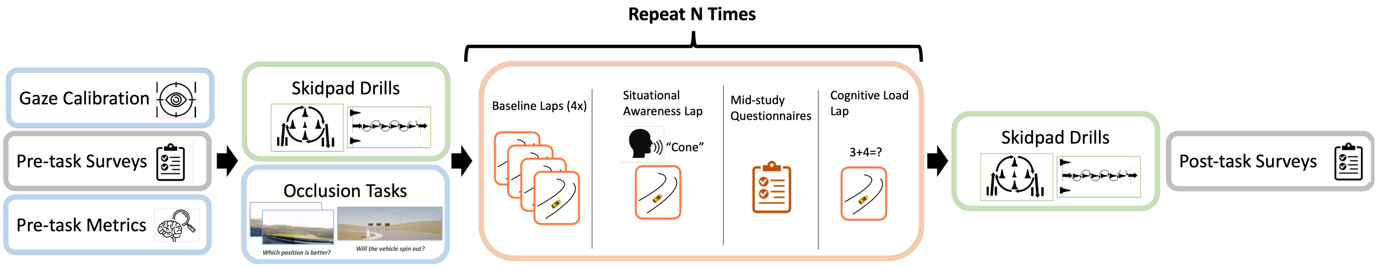

Fig. 2: This figure outlines our study procedure. Participants start with pre-task assessments, surveys on HPDE experience, stress performance, and personality. After gaze calibration, they complete skidpad drills and occlusion tasks, then alternate between lap driving and questionnaires. The session ends with repeated skidpad drills and a final questionnaire.

- Mental Skill: Changes in fatigue, cognitive load, and psychological resilience and regulation [16].
- Gaze Policy: Ability to visually focus and look ahead while driving [17], based on eye tracking [18]
- Vehicle Handling: racing line distance, ability to handle vehicle at the limits, and performance on skidpad drills.
- Perceptual: performance on occlusion tasks [19].

We note that consistency is also a core component of skill [1]. Skilled drivers demonstrate consistency by performing each of the above subskills with a high degree of reliability over time. [1].

D. Data Collection and Dataset

Given our insights from literature and coaches for how to assess and measure skill in HPDE, we next collected a dataset to train our AI system to extract skill representations. The dataset comprises both task data (i.e., trajectory data from students driving laps around the track) and skill metrics aimed at assessing various aspects of skill. We recruited 18 participants from Toyota Research Institute, including 9 novice and 9 experienced drivers in HPDE (age: 55% 26-35, 38% 36-45 and 5% over 45; 16% Female). We define experience drivers as participants who have experience in performance and are in the bottom quartile of lap times. We define novices as participants who have never participated in HPDE and are in the top quartile of lap times.

In our data collection study, participants complete tasks and questionnaires to evaluate their know-how, reaction time, hand-eye coordination, grip strength, and previous HPDE experience. Next, participants enter the compact simulator (see Appendix A for details), calibrate gaze-tracking [18], and proceed with occlusion tasks and skidpad drills.

Participants next engage in repeated blocks, each consisting of four laps of regular driving around a track, a situational awareness lap, a questionnaire, and a final cognitive load lap. In the situational awareness lap, the participant is tasked with saying "cone" as soon as they see a cone on the side of the track to assess their awareness of their surroundings while driving. The questionnaires evaluates mental fatigue, perceived workload, enjoyment, emotional state, and level of motion sickness. In the cognitive load lap, participants complete the secondary task of answering math questions while driving.

These tasks and metrics are repeated until approximately 15 minutes remain, with participants typically completing 3-4 blocks. The goal of repeating blocks is to assess consistency and identify predictors of performance changes. In the remaining time, participants revisit skidpad drills and complete a final questionnaire. Participants wear the Empatica device to collect physiological measures throughout the study. More detailed descriptions of each task and metric are in Appendix D.

E. Data-Driven Skill Representation

The ultimate goal of this work is to leverage the collected data to train a machine learning model capable of extracting a meaningful representation of skill from simple trajectory data, τ (e.g., position and control inputs). Figure 1 illustrates a high-level overview of the proposed architecture, detailing how each component interacts. We aim to learn two distinct representations from the task data: a performance representation (denoted as z^p) and a skill representation (denoted as z^s). These representations not only serve as the foundation for understanding driver behavior but will also be used for inference by an AI system, enabling downstream tasks such as personalized coaching and adaptive assistance. In the following sections, we present an overview of our modeling objectives but leave the details and implementation to future work.

Performance Representation (z^p): The representation of performance is primarily driven by the output of driving laps around the track, with lap time and other performance-related metrics such as smoothness, sector times, and speed at important locations as the key supervision signals. Lap time is a well-established metric for evaluating driving performance and serves as a direct indicator of how well the driver is executing the task.

Additionally, we incorporate input factors that may influence or predict fluctuations in performance, such as fatigue, cognitive load, and motion sickness. These factors help distinguish between momentary performance variations and underlying skill, enabling the model to better infer declines in performance due to transient conditions rather than a lack of ability. By learning a representation of performance, the system can adapt coaching strategies dynamically, provide targeted interventions, or prompt further queries to diagnose the cause of performance variability.

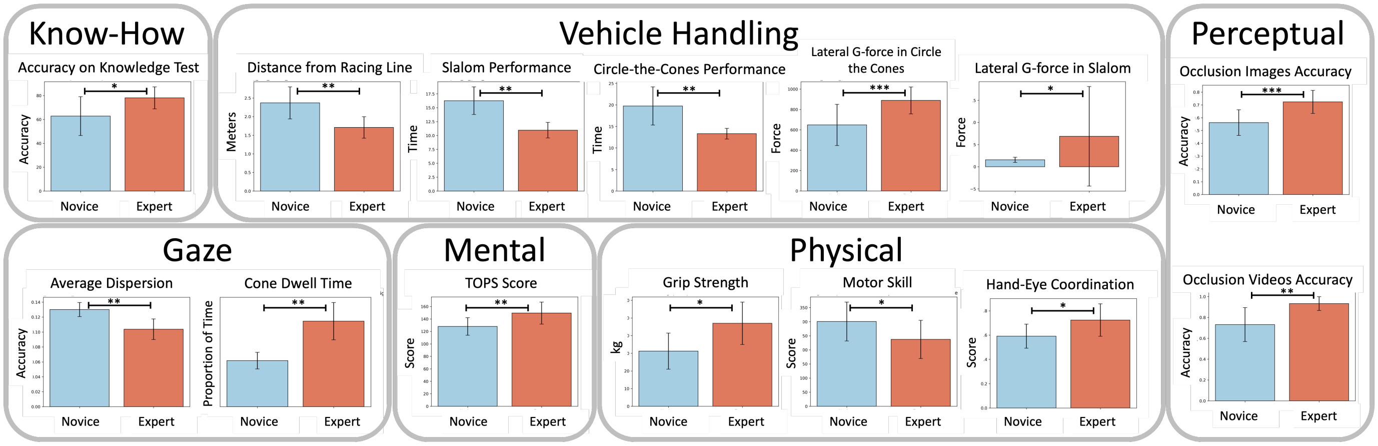

Fig. 3: This figure shows that novice and experienced drivers exhibit significant differences across a set of skill metrics.

Skill Representation (z^s): In contrast, the skill representation captures a driver's underlying ability. We aim to extract this skill representation from trajectory data by using skill-specific metrics collected in this work. These metrics act as the supervision signal for learning z^s, guiding the model to distinguish skill-related aspects of the driving behavior from mere performance outcomes. We also aim to establish a prior on an individual driver's skill, z_0^s that enables the AI system to begin to personalize its interactions before it has even observed the user driving, thus avoiding the cold-start problem [20]. We propose to learn this skill prior via pre-task metrics collected prior to driving. During inference, z^s will be used to estimate a driver's skill level over time, supporting longitudinal tracking and enabling skill-adaptive coaching and assistive strategies.

Interaction Between Skill and Performance: The interaction between skill and performance is central to the proposed architecture. Grounded in existing literature, we model performance and skill as distinct yet interconnected constructs. The model is designed to disentangle these influences by using two distinct representations—for performance and for skill—yet it also accounts for how skill influences performance over time. Furthermore, we leverage the skill and performance embeddings to learn to reconstruct the input trajectory to ensure that latent space preserves essential task information.

IV. Preliminary Results

Given our dataset of skill metrics from novice and experienced drivers in HPDE, our first goal is to verify our hypothesis: that these metrics can effectively distinguish between experienced and novice drivers and therefore provide useful supervision signals to our AI system. To test our hypothesis, we conducted a t-test if the data met parametric assumptions and a Wilcoxon test otherwise. The details are reported in Appendix H. Figure 3 shows significant differences between novices and experts across all skill metrics, confirming their effectiveness in distinguishing skill levels.

Specifically, for **know-how** subskill, experienced drivers outperform novices on the knowledge test (p = .045, d = $-.85$). For the **vehicle handling** subskill, expert drivers demonstrate greater accuracy in following the racing line (p = .005, d = -1.71). On the quick skidpad drills expert drivers show greater proficiency (p = .004, d = .9 and p = .004, d = .8) and are better able to push their vehicle to the limits (p = .019, d = $-.98$ and p = .0004, d = -1.95).

Additionally, we hypothesized that expert drivers would exhibit more focused **gaze**. Our results confirm this, with experts showing significantly more concentrated gaze patterns (p = .0003, d = 2.1) and spend significantly more time focusing on the cornering cones (p = .006, d = $-.97$). In terms of **mental** skill, novice drivers report significantly lower scores on the TOPS psychological resilience survey [16] (p = .008, = -1.26).

In the **physical** skill domain, experts show superior grip strength (p = .02, d = -1.1), motor skill (p = .044, d = .85), and hand-eye coordination (p = .042, d = $-.93$). For **perceptual skill** expert drivers perform significantly better on the occlusion tasks (p = .0001, d = -2.2 and p = .009, d = -1.2), further reinforcing the distinction between skill levels. **These findings provide evidence that our skill metrics effectively differentiate skill levels, supporting their use as supervision signals for AI-driven skill modeling.**

V. Future Work

Future work will expand the dataset to include a broader range of skill levels, refine supervision signals for accurate representations, and investigate how sub-skill differences manifest themselves in the trajectory data. We then aim to evaluate our approach's ability to accurately capture skill in HPDE.

Validation will focus on four key aspects: (1) *Predictive Validity*—can the model predict relevant outcomes? (2) *Expert Alignment*—does it match how coaches assess skill? (3) *Construct Validity [21]*—is it stable over time and reflective of learning progress? and (4) *Utility*—does it enhance real-world applications like personalized coaching? While we focus on HPDE, future work will probe whether the learned skill representation generalizes to other long-term HRI applications, such as adaptive robotic assistance, personalized coaching in sports, and driver assistance systems.

REFERENCES

[1] Seppo E. Iso-Ahola. A theory of the skill-performance relationship. *Frontiers in Psychology*, 15, 2024.

[2] Xin Zhang, Ruiqing Zhang, Deqing Huang, and Yanan Li. Robot skill learning and generalization based on human-robot collaboration and computer vision. In *2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS)*, pages 2095–2100, 2024.

[3] Ghodai Abdelrahman, Qing Wang, and Bernardo Nunes. Knowledge tracing: A survey. *ACM Comput. Surv.*, 55(11):1–37, February 2023.

[4] Montana Office of Public Instruction. *Behind-the-Wheel Instruction Guide*, 2017. Accessed: 2025-02-03.

[5] Daniela Ndunguru, Liu Zhanwen, Chrispus Zacharia Oroni, Seth Mabyo Kabamba, Arsenyan Ani, Moussa Sali, Gadi Gilleard Lyatuu, and Aletas Athanas Haule. Predictive modeling of gaze patterns in drivers. *International Journal of Machine Learning and Cybernetics*, XX(YY):ZZ, 2024. Accessed: 2025-02-03.

[6] Mariah L Schrum, Erin Hedlund-Botti, and Matthew C Gomoblay. Reciprocal mind meld: Improving learning from demonstration via personalized, reciprocal teaching. *Conference on Robot Learning*, 2022.

[7] Deepak Gopinath, Xiongyi Cui, Jonathan DeCastro, Emily Sumner, Jean Costa, Hiroshi Yasuda, Allison Morgan, Laporsha Dees, Sheryl Chau, John Leonard, et al. Computational teaching for driving via multi-task imitation learning. *arXiv preprint arXiv:2410.01608*, 2024.

[8] Yuchen Cui Deepak Gopinath Emily Sarah Sumner Andrew Silva Laporsha Dees Guy Rosman Dorsa Sadigh Megha Srivastava, Reihaneh Iranmanesh. Shared autonomy for proximal teaching. In *HRI*, 2025.

[9] Kyle Lam, Junhong Chen, Zeyu Wang, Fahad M. Iqbal, Ara Darzi, Benny Lo, Sanjay Purkayastha, and James M. Kinross. Machine learning for technical skill assessment in surgery: a systematic review. *npj Digital Medicine*, 5(1):24, 2022.

[10] Chunsheng Yang, Feng-Kuang Chiang, Qiangqiang Cheng, and Jun Ji. Machine learning-based student modeling methodology for intelligent tutoring systems. *Journal of Educational Computing Research*, 59(6):1015–1035, 2021.

[11] Richard A. Magill and David I. Anderson. McGraw-Hill Education, New York, NY, 2018.

[12] Pravin Chopade, David Edwards, Saad M. Khan, Alejandro Andrade, and Scott Pu. CPSX: Using AI-machine learning for mapping human-human interaction and measurement of CPS teamwork skills. In *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*, pages 1–6, 2019.

[13] Juan M. García-Ceberino, María G. Gamero, Sebastián Feu, and Sergio J. Ibáñez. Experience as a determinant of declarative and procedural knowledge in school football. *International Journal of Environmental Research and Public Health*, 17(3), 2020.

[14] J Backman, K Häkkinen, J Ylinen, A Häkkinen, and H Kyröläinen. Neuromuscular performance characteristics of open-wheel and rally drivers. *Journal of Strength and Conditioning Research*, 19(4):777–784, November 2005.

[15] I. Jukic, K. Prnjak, A. Zoellner, J. J. Tufano, D. Sekulic, and S. Salaj. The importance of fundamental motor skills in identifying differences in performance levels of u10 soccer players. *Sports (Basel)*, 7(7):178, July 2019.

[16] A. J. Morgan. *An Examination of the Psychological Skills Profiles of Oval Racers and Road Racers*. Ph.d. dissertation, ProQuest Dissertations Theses AI, 2013. Order No. 1546320. Available from ProQuest Central; ProQuest One Academic (1450049062).

[17] Peter M. van Leeuwen, Stefan de Groot, Riender Happee, and Joost C. F. de Winter. Differences between racing and non-racing drivers: A simulator study using eye-tracking. *PLOS ONE*, 12(11):1–19, 11 2017.

[18] Tobii. *Tobii Spark Pro User Guide*, 1.2 edition, 2024. Available at https://example.com/manual.pdf.

[19] Fran Allard and Janet L. Starkes. Perception in sport: Volleyball. *Journal of Sport Psychology*, 2(1):22 – 33, 1980.

[20] M. Brent Donnellan, Frederick L. Oswald, Brendan M. Baird, and Richard E. Lucas. The mini-ipip scales: Tiny-yet-effective measures of the big five factors of personality. *Psychological Assessment*, 18(2):192–203, 2006.

[21] Caroline Wehner, Lena Roemer, and Matthias Ziegler. *Construct Validity*, pages 875–878. Springer International Publishing, Cham, 2020.

[22] Gijsbert Stoet. Psytoolkit: A software package for programming psychological experiments using linux. *Behavior Research Methods*, 42(4):1096–1104, 2010.

[23] Paul M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6):381–391, 1954.

[24] Josh de Leeuw. jspsych serial reaction time plugin, 2023.

[25] David Watson, Lee Anna Clark, and Auke Tellegen. Development and validation of brief measures of positive and negative affect: The panas scales. *Journal of Personality and Social Psychology*, 54(6):1063–1070, 1988.

[26] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in Psychology*, volume 52, pages 139–183. Elsevier, 1988.

[27] Jason E. George. *Maximizing the Development of Tactical Decision-Making in Perceptual-Motor Skills with Athletes*. PhD thesis, ProQuest Dissertations and Theses, 2022.

[28] S. Müller, K. Morris-Binelli, D. Z. Hambrick, and B. N. Macnamara. Accelerating visual anticipation in sport through temporal occlusion training: A meta-analysis. *Sports Medicine*, 54(10):2597–2606, October 2024. Epub 2024 Aug 5.

[29] Ying Shuai, Shaoshen Wang, Xian Liu, Yee Cheng Kueh, and Garry Kuan. The influence of the five-factor model of personality on performance in competitive sports: a review. *Frontiers in Psychology*, 14, 2023.

[30] Patricia Laguna Melinda Frey and Kenneth Ravizza. Collegiate athletes' mental skill use and perceptions of success: An exploration of the practice and competition settings. *Journal of Applied Sport Psychology*, 15(2):115–128, 2003.

[31] Takaaki Kato. Visual behaviors and expertise in race driving situation. In Sebastiano Bagnara, Riccardo Tartaglia, Sara Albolino, Thomas Alexander, and Yushi Fujita, editors, *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, pages 245–248, Cham, 2019. Springer International Publishing.

[32] Joan N. Vickers. Mind over muscle: the role of gaze control, spatial cognition, and the quiet eye in motor expertise. *Cognitive Processing*, 12(3):219–222, 2011.

[33] Bruce Abernethy and David G. Russell. Expert-novice differences in an applied selective attention task. *Journal of Sport Psychology*, 9(4):326–345, 1987.

[34] Bruce Abernethy. Anticipation in squash: Differences in advance cue utilization between expert and novice players. *Journal of Sports Sciences*, 8(1):17–34, 1990.

[35] Donghyun Ryu, Bruce Abernethy, David L. Mann, Jamie M. Poolton, and Adam D. Gorman. The role of central and peripheral vision in expert decision making. *Journal of Sport and Exercise Psychology*, 35(3):265–277, 2013.

APPENDIX

A. Pre-Study Survey Metrics

These metrics were gathered prior to the driving sessions to understand initial skill and experience levels. Their selection was informed by interviews with experienced HPDE coaches and a review of relevant literature.

- **Hand-Eye Coordination Task** - Assesses a participant's ability to coordinate precise motor movements in response to changing dynamics and can act as an indicator of driving proficiency. We specifically evaluate hand-eye coordination via Fitts' law coordination on the Fitts task [22], [23].
- **Motor Skill** - Assesses a participant's fundamental movement execution and control, independent of visual guidance. Unlike hand-eye coordination, which involves integrating visual input with motor responses, motor skill reflects the efficiency and precision of movement itself. We evaluate motor skill using the average response time on the Fitts task [22], [23].

- **Serial Reaction Time Test** - Assesses participants' ability to respond quickly to stimuli, correlating with reaction speed in driving. To measure this, we leveraged JsPsych's serial reaction time task [24].
- **Demographics** - Includes participant age, gender, prior HPDE experience, years of driving, and frequency of track racing.
- **PANAS (Positive and Negative Affect Schedule - Short Form)** - Evaluates participants' emotional state before driving to examine potential effects on driving performance.
- **Short Stress Questionnaire** - Measures baseline stress levels before driving to assess their impact on cognitive load and performance, enabling comparisons with post-driving stress responses.
- **Fatigue** - Measures baseline fatigue levels before driving to assess their impact on cognitive load and performance, enabling comparisons with post-driving fatigue levels.
- **Big 5 Personality Test** - Measures core personality traits to analyze how personality influences driving behavior and adaptability [20].
- **Youth Experience with Sports & Current Sports Experiences** - Identifies past and present engagement in sports to assess potential motor learning transfer effects.
- **Driving Type and History** - Captures self-reported driving habits, including preferred driving conditions and history of aggressive/defensive driving styles.
- **TOPS-2 (Test of Performance Strategies - 2nd Edition)** - Evaluates psychological skills and coping strategies relevant to high-performance driving [16].
- **Self-Assessment of Skill** - Participants rate their own driving skill to compare with objective skill measures.
- **Racing Knowledge Assessment** - A multiple-choice test covering racing line principles, vehicle dynamics, and track safety.

B. Mid-Study Survey Metrics

Collected at regular intervals to track cognitive and physiological changes throughout the driving task.

- **PANAS (Repeated Post-Lap Evaluation)** - Measures changes in affect during the study to assess emotional fluctuations [25].
- **Fast Motion Sickness Questionnaire (FMSQ)** - Assesses symptoms of motion sickness, which can impact driving performance and concentration.
- **Performance and Satisfaction Rating** - Participants rate their own performance and satisfaction with their driving after each set of laps.
- **Self-Reported Skill Assessment** - Assesses whether participants perceive improvements in their driving ability.
- **Self-Reported Track Familiarity** – Assesses participants' familiarity with the track over the course of the study.
- **NASA TLX (Task Load Index)** - Measures perceived workload across physical, mental, and temporal demand dimensions [26].
- **Enjoyment Scale** - Captures enjoyment levels to determine the impact of task engagement on skill acquisition.

C. Post-Study Survey Metrics

Administered after participants completed all driving tasks to assess overall experience and perception.

- **PANAS (Final Assessment)** - Captures emotional state at the end of the study for comparison with initial PANAS scores.
- **Stereotype Threat Questionnaire** - Examines whether participants felt pressure based on preconceived notions about driving skill.
- **General Feedback & Open-Ended Questions** - Allows participants to provide qualitative feedback about the driving task and AI coaching.

D. Skill Metrics

These metrics were extracted from driving task data to evaluate participants' skill. Each subskill is described in detail, along with the specific measures used.

1) Know-How:

- **Racing Knowledge Test** - A multiple-choice test covering track rules, racing lines, and vehicle dynamics to assess theoretical understanding.

2) Physical:

- **Grip Strength Test** - Assessed using a hand dynamometer to measure participants' ability to maintain precise vehicle control.
- **Reaction Time Assessment** - Measures the time taken to respond to visual cues, correlating with motor response effectiveness.
- **Hand-Eye Coordination Task** - Evaluates a participant's ability to coordinate visual input with precise motor actions, which is critical for steering precision.

3) Mental:

- **Cognitive Load** - Involves a dual-task assessment (e.g., solving math problems while driving) to measure the ability to maintain performance under cognitive strain, providing key insights into skill differentiation.
- **Situational Awareness** - Participants must identify trackside objects (e.g., cones) while driving to assess their awareness levels.
- **Stress** - Collected via Empatica device to measure heart rate variability and skin conductance. This measurement aims to determine whether emotional regulation can serve as a predictor of skill and influence performance outcomes.
- **Mental Fatigue** - Collected via questionnaires during every block to assess changes in mental fatigue
- **Psychological Resilience** - captured via the TOPS-2 questionnaire [16].

4) Gaze Policy:

- **Gaze Dispersion** - Measures the spread of a driver's visual attention. We hypothesize that experts distribute their gaze more efficiently, scanning ahead to prepare for future actions, while novices focus more on immediate obstacles.

- **Gaze Position** - Captures the average location of a driver's visual focus during a driving task. We hypothesize that experienced drivers will tend to look further ahead.
- **Gaze Fixation Points** - Identifies specific locations where the driver fixates and measures fixation duration, frequency, and transitions. We hypothesize that experienced drivers will prioritize critical track features (e.g., apexes, braking points).

5) Vehicle Handling:

- **Quick Skidpad Drills** - Designed to assess vehicle handling and car control at the limits of grip. These include a slalom task, where drivers weave between a series of cones to evaluate steering precision and weight transfer control, and a circle-the-cones task, where drivers navigate a circular path around cones to test their ability to maintain consistent lateral G-forces and execute smooth throttle and steering inputs.
- **Lateral G-Force on quick skidpad drills** - Measured by calculating the average lateral g-force of the vehicle
- **Steering Smoothness** - Measured by analyzing micro-adjustments in steering inputs.
- **Spinouts** - Number of times that the participant looses control of the vehicle.
- **Deviation from Optimal Racing Line** - Measured using trajectory data to evaluate how well participants follow an ideal path.
- **Braking Consistency** - Evaluates braking force application and consistency in cornering scenarios.
- **Acceleration Control** - Assesses throttle application and transition smoothness between braking and acceleration phases.

6) Perceptual:

- **Occlusion Image Task** - Participants are given half a second to choose between two images taken in a first-person-point of view and the participant must determine which vehicle is position better on the track. This task aims to examine whether experienced drivers can accurately assess and predict outcomes via limited visual cues.
- **Occlusion Video Task** - Participants are shown video clips of drivers approaching key points on the track, some successfully completing the maneuver and others not. The video is paused before the maneuver completes. The participant is then asked to predict if the vehicle will succeed in the maneuver. This task aims to assess the participant's ability to leverage the visual and auditory cues to make accurate judgements and predictions.

E. Interview Guide with HPDE Coaches

The following questions were used in our semi-structured interviews with expert HPDE coaches to gain insights into skill assessment and representation:

1) Definitions:

- How do you define performance? How do you define skill?

2) Prior on Skill:

- Can you walk me through your interactions with a student prior to them hitting the throttle? What questions do you ask them and what do you pay attention to?
- What does this information tell you?
- What are inherent traits, experiences (apart from driving), and/or characteristics of a student that are informative to you as a coach?
- What does this information tell you and how may it influence your evaluation of their potential?
- How much can a student's skill improve with self-practice?
- How much self-practice is required to see a significant improvement in skill, if this is even possible?
- What characteristics of a student inform how their skill might evolve over time?
- If you are meeting a student for the first time, do you ask questions not directly related to driving that help you predict the student's initial abilities?
- Is there any demographic information that helps you predict the student's initial abilities or that influences your coaching?
- How important is the student's previous racing or driving experience to your coaching approach?
- What specifically about their prior experience is predictive?
- Do you try to understand a student's personality traits?
- If so, are there any traits that you find indicative of competence or that provide insight into a participant's initial skill level?
- To what extent do personality traits have predictive power in skill development, and are these traits considered more important at the initial stages of learning, or do they become more influential as learning progresses?
- Do you try to understand a student's confidence levels or resilience under pressure?
- Do you ask about a student's physical fitness?
- Are there aspects of physical fitness that can be indicative of skill development?
- Do you ask if a student has previous experience with sports?
- If so, why and what does that information reveal to you in relation to the student's potential skill development?
- Does familiarity or exposure to other sports provide additional advantages to skill building in this context?

3) Subskills:

- What are the sub-skills (physical or psychological) that are essential to mastery in high-performance driving?
- How do you evaluate each of these subskills?

4) Evaluating Skill:

- How do you as a coach evaluate a student's skill?

- How much time is required to make a confident evaluation?
- How much time/number of laps is required to see a significant change in skill?
- What are specific metrics that you use to evaluate skill?
- What additional information besides these metrics helps to inform your evaluation?
- Why do you use these specific metrics?
- What are characteristics of a student's driving that give you insight into their skill?
- What are driving characteristics that differentiate an expert from an intermediate from a novice in terms of skill?
- Are there any tasks or drills that help you evaluate a student's skill apart from doing laps around the track?
- What are skill metrics that you would use in these tasks?
- Are there any tasks or drills or areas of focus that help you evaluate a student's skill while doing laps around the track?

5) Variables Predictive of Performance:

- Before each racing lap, what questions would you ask a student that would help you predict how well they will perform in the next lap or that might help you explain their performance in the previous lap?
- What factors might affect a student's performance from lap to lap?
- If you notice a student's performance is suffering/different from what you'd expect, do you ask them questions to understand why?
- Could the mental state of the student impact their performance?

F. Key Insights From Interviews

1) Know-How:

> "I ask a lot of different questions—ranging from any experience doing indoor karting, driving simulators, or video games. That tells me if they are comfortable with speed and track environments before they even get behind the wheel." - C1

> "If they've done coaching before or if they've kind of dived into reading about the concepts of this, and how to put these things in practice, they're going to do better. They have a better foundation of an understanding of what they're trying to achieve." - C3

2) Physical:

> "I'd say cardiovascular health is quite [important], depending on if we're talking wheel-to-wheel or a high-stress environment. Like cardiovascular health is probably the biggest thing." - C3

> "Depending on the type of car that we're driving, it could also be muscle health, right? Because how we're operating the vehicle takes strength and stamina. There's a lot of cars that have no power assist on them. So, you're going to need to have kind of that muscle health to continue to operate the car at a high level for hours on end if that's how you're racing." - C3

> "We use drills for hand-eye coordination a lot in high-level motorsport, like differentiating colors on button pads or using reaction time tasks." - C3

3) Gaze Policy:

> "A novice is going to look down the nose of the car or look maybe 50 feet ahead of the vehicle, and that's just not enough information for us to achieve the goals that we're looking for. An expert is scanning corner towers, scanning mirrors, looking at the apex and exits every single time." - C3

> "I tell them all to wave at flaggers so they know exactly where they are. If they wave at four out of five, then I know that they're not looking in the proper spot." - C1

> "The car will do exactly what you tell it to do, but you need to understand that if you don't look ahead, you'll be late to every input." - C3

4) Vehicle Handling:

> "The first thing I do is assess how a driver is working with the vehicle. How are their inputs going into the steering wheel, brake pedal, throttle? Are they smooth? Are they reacting properly to the car?" - C3

> "If someone is really sawing at the wheel, they're not in control. You should be able to turn the wheel, place the car where it needs to go, and straighten the wheel." - C1

> "The very basic catch and release that I was talking about, that very first fundamental thing is I will send them out. Here's a single cone in the center of the skid pad. Make the car slide around that cone. That's going to tell me almost immediately where we're at." - C3

> "On a skid pad, if you just get them sideways, how do they do it? Are they frantically grabbing the wheel and turning it? Are they calm, collected, and when the back end comes out, they put their eyes in the proper place?" - C1

5) Perceptual:

> "Perceptual-motor skills in sports are competencies that combine processing sensory-information that is identified in athletic settings and coordinating it with trained physical movements" - [27]

> "These findings are evidence that the temporal occlusion paradigm is an effective method to improve visual anticipation skill across representative perceptual and perceptual-motor transfer tests." - [28]

6) Mental Skill:

> "Novices don't have the mental capacity to process multiple things at once—other cars, flag stations, braking points. If they have to process two things at once, something suffers." - C3

> "An ideal driver would be able to control their heart rate, not panic, and not get frustrated or overstimulated. Just be patient and calm." - C1

G. Description of Driving Simulator

Our study utilized a compact driving simulator to evaluate participants' driving skills in a controlled and repeatable environment. The simulator setup and capabilities are described below:

1) Hardware Specifications (Fig. 4):

- **Simulator Frame** - A stationary driving rig with adjustable seating to accommodate different driver body types.

Fig. 4: This figure shows our driving simulator.

– **Steering System** - A high-fidelity force-feedback Fanatec steering wheel providing realistic resistance and road feel.
– **Pedal System** - A pressure-sensitive Fanatec pedal set with adjustable resistance to simulate real-world braking and acceleration dynamics.
– **Monitor Mount** - A Trak Racer mount system supporting the display to enhance stability and realism.
– **Seating** - An integrated racing-style seat to mimic real driving posture and comfort.

2) *Software Environment:*

– **Simulation Software** - The simulator runs on CARLA, an advanced open-source driving simulator used for research purposes.
– **Physics Engine** - CARLA provides high-fidelity vehicle physics, simulating realistic tire grip, weight transfer, and aerodynamics.
– **Car and Track Models** - The simulator features a race-car dynamics profile, ensuring that vehicle inputs and outputs closely resemble real-world race conditions.
– **Track Environment** - The driving scenarios take place on a digital recreation of Thunderhill Raceway, accurately modeled for realism.

H. Literature Review

To identify skill metrics that differentiate skill levels, we review the high-performance driving literature and examine broader sports research to explore transferable insights.

– Big Five Personality - Prior research suggests that personality influences sports-related abilities. Specifically, neuroticism is negatively correlated with sports performance, while extraversion and conscientiousness are positively associated with success [29].
– Motor Skills and Hand Eye - Studies indicate that motor skills are predictive of sports performance, particularly among youth athletes [15].
– Mental Skill in Sports - Literature suggests that a strong relationship exists between mental skills (e.g., handling pressure, maintaining a positive mindset) and success across various sports [30]. Research in motorsports specifically has found a positive correlation between mental skills and racing performance [16].
– Gaze - Prior research provides strong evidence that experienced drivers exhibit distinct gaze fixation patterns compared to novices. Experts tend to focus on specific points, particularly while turning, and look further ahead on the track [17], [31], [32].
– Grip Strength - Prior studies report that experienced high-performance drivers exhibit greater grip strength, with rally drivers demonstrating stronger grip than open-wheel racers [14].
– Occlusion tasks - Seminal work by Allard and Starkes [19] introduced the paradigm of the occlusion task to examine how expertise influences visual processing in sports. Their study on volleyball players demonstrated that experts recognize meaningful patterns and anticipate outcomes more effectively than novices, even when key visual elements are removed or only displayed for a short period of time. Follow-up research explored both temporal and spatial occlusion tasks. In Abernathy and Russel [33], participants were shown video clips of an opponent performing a movement (e.g., a tennis stroke), with the footage occluded at various points—before, during, or after critical movement cues. The study found that experts were significantly better at predicting outcomes based on early visual information, indicating more advanced perceptual-cognitive skills. Since then, occlusion tasks have been widely used to distinguish novices from experts across various sports. [34], [35].

Metric	**Test**	**P-Value**	**Statistic**	**Normality Novice**	**Normality Expert**	**Homoscedasticity**
Knowledge Test	t-test	.045	t(16)=-1.8	.12	.14	.84
Distance From Racing Line	Wilcoxon	.005	W=44	.56	.046	.67
Slalom Performance	t-test	.004	W=45	.51	.018	.093
Circle-the-Cone Performance	Wilcoxon	.004	W=45	.48	.99	.002
Slalom G-Force	Wilcoxon	.019	W=5	.38	.001	.29
Circle-the-Cone G-Force	t-test	.0004	t(16)=-4.13	.87	.20	.80
Average Dispersion	t-test	t(16)=4.32	.0003	.7	.31-	-.05
Cone Dwell time	Wilcoxon	.006	W=0	.18	.21	.006
TOPS Score	t-test	.008	t(16)=-2.68	.10	.30	.42
Grip Strength	t test	.02	t(16)=-2.2	.88	.33	.89
Motor Skill	t-test	.044	t(16)=1.81	.611	.54	.53
Hand Eye Coordination	Wilcoxon	.042	W=8.	.046	.093	.3
Occlusion Image Task	t-test	.0001	t(16)=-4.7	.48	.26	.21
Occlusion Video Task	t-test	.009	t916)=-2.6	.34	.12	.055

TABLE I: Results of statistical tests. For each metric, if the data did not pass parametric assumptions (normality and homoscedasticity, we conducted a Wilcoxon tet. Otherwise we conducted a t-test. We report the p-value and test stastic in the above table.