JAIN
RESEARCH LAB

Online Self-Confidence Calibration for Improving Learning Outcomes via Intelligent Tutoring Systems

Research Assistant: Madeleine Yuh (myuh@purdue.edu) | Principal Investigator: Dr. Neera Jain (neerajain@purdue.edu)

PURDUE UNIVERSITY

Ray W. Herrick Laboratories

OVERVIEW

Intelligent tutoring systems (ITS) are used to train humans by personalizing education systems. For conventional learning contexts such as mathematics, agents in ITSs have been designed to respond to humans based on cognitive feedback such as self-confidence and workload. However, the same cannot be said for psycho–motor learning contexts. Existing psychomotor ITSs face the following challenges:

1. Creating a knowledge space of the task.
2. Maintaining learner motivation to learn.
3. Personalizing the agent to the learner's characteristics.

So far, the first challenge has been addressed through the development of an online learning stage classifier, bridging the gap between qualitative and quantitative representation of learning stage (novice, advanced beginner, competent, proficient, and expert) [1]. The second challenge has been addressed through development of an optimal cognitive control policy trained using RL (reinforcement learning) methods to determine when to provide assistance to learners [2].

By leveraging tools developed in prior work, the **main objective of this work is to evaluate whether using automation that assists learners based on an algorithm designed to calibrate self-confidence to performance leads to improved learning outcomes in comparison to learners receiving no assistance** [3].

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Award No.183690. Any opinions, findings, and material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] M. S. Yuh, K. R. Ortiz, K. S. Sommer-Kohrt, M. Oishi and N. Jain, "Classification of Human Learning Stages via Kernel Distribution Embeddings," 2024.

[2] M. Yuh, E. Rabb, A. Thorpe, and N. Jain, "Using Reward Shaping to Train Cognitive–based Control Policies for Intelligent Tutoring Systems," in Proceedings of the American Control Conference, Toronto, ON, CANADA, July 10-12, 2024.

[3] M. S. Yuh and N. Jain, "Online Self-Confidence Calibration for Improving Learning Outcomes Via Intelligent Tutoring Systems", in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Phoenix, AZ, September 9-13, 2024.

Figure 1: *in-person experimental platform.*

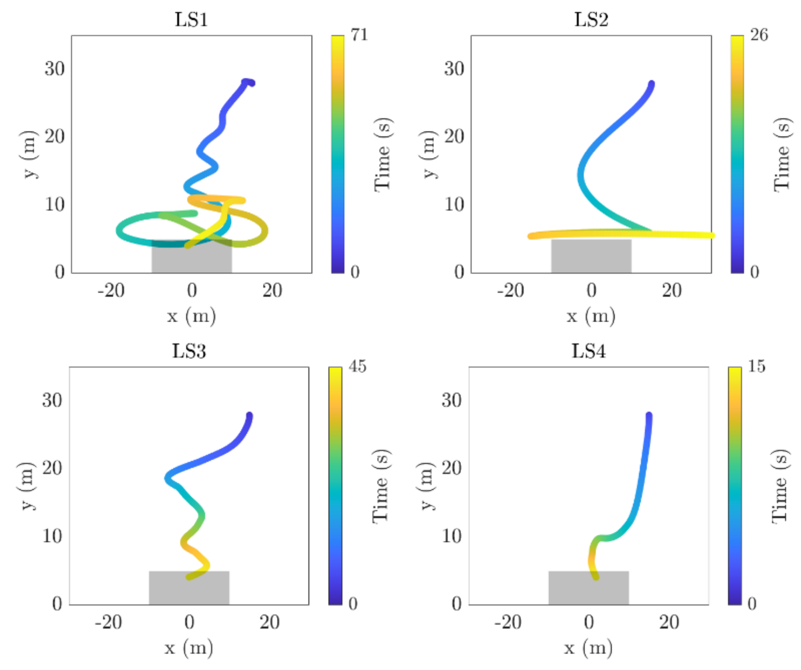

Figure 2: *Example Quadrotor Trajectories for each learning stage.*

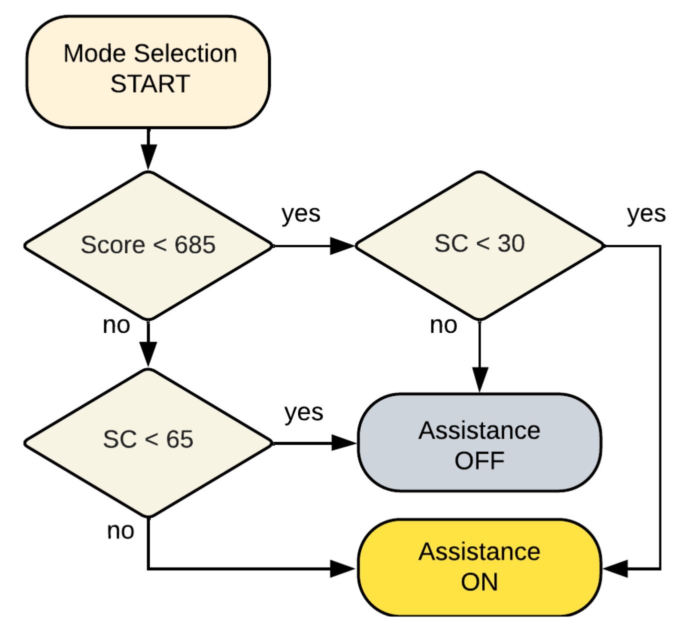

Figure 3: *Optimal Control Policy*

Experiment Design

- **Participant's Goal:** Learn to manually land quadrotor safely onto landing pad within 25 trials.
- Cognitive control algorithm chooses when to assist user (Figure 3).
- After every trial k, numerical score ($S_k \in [0,1000]$, self-reported self-confidence ($SC_k \in [0,100]$), self-reported mental workload ($W \in [0,100]$), assigned control mode (shared, manual), landing type (unsuccessful, unsafe, safe), quadrotor states ($x, y, \phi, \dot{x}, \dot{y}, \dot{\phi}$) and classified learning stage ($LS_k \in \{1,2,3,4\}$) are collected.
- The Institutional Review Board at Purdue University approved the study. Participants are compensated at $20/hr.
- 24 participants split evenly into two groups (**Algorithm Group:** Assisted by algorithm, **Manual Group:** Complete all trials manually)

Results

After collecting data, we first utilize independent t-tests to compare self-reported data, performance metrics, and achieved learning stages shown in **Table 1**. Then, we complete a multivariate regression analysis on self-confidence using the feedback information given to learners after and during each trial shown in **Tables 3-5**.

*Table 1: Independent t-test results between the algorithm and manual groups including t-value, degrees of freedom (DOF), p-value, significance (sign.), means, and standard deviations. Significant variables are in **bold**.*

					Algorithm Group		Manual Group	
Dependent Variable	t-value	DOF	p-value	Sign.	Mean	Standard Deviation	Mean	Standard Deviation
Self-confidence in trials 21-25	0.134	118	0.894		80.8	20.7	80.2	26.6
Mental Workload in trials 21-25	-2.05	118	0.0424	*	37.0	19.1	44.8	22.7
Performance scores in trials 21-25	0.0802	118	0.936		867	183	864	174
Unsuccessful landings in trials 21-25	0.196	22	0.847		0.833	1.11	0.750	0.965
Unsafe landings in trials 21-25	1.25	22	0.223		0.667	0.779	0.333	0.492
Safe landings in trials 21-25	-0.758	22	0.457		3.50	1.51	3.92	1.16
LS1 classifications in trials 21-25	-0.158	22	0.876		0.917	1.38	1.00	1.21
LS2 classifications in trials 21-25	-1.23	22	0.233		1.42	1.93	2.33	1.72
LS3 and LS4 classifications in trials 21-25	1.31	22	0.205		2.67	1.78	1.67	1.97
Total LS1 classifications	-0.668	22	0.511		6.75	5.46	8.08	4.23
Total LS2 classifications	-2.09	22	0.0489	*	7.58	5.84	12.3	5.10
Total LS3 and LS4 classifications	2.87	22	0.00852	**	10.7	6.02	4.67	3.94

Note: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Part I: Independent *t*-tests

- Difference in means of scores, self-confidence, number of LS1-LS4 classifications, and number of landing types in trials 21-25 are **not significant** between groups.
- Mean mental workload in last five trials is **significantly lower** for the algorithm group**.**
- Mean total LS2 classifications and mean total LS3 and LS4 classifications are **significantly lower and higher, respectively**, for the algorithm group due to receiving assistance that augments the user's input to be more like that of an expert's input.

Table 3: *Coefficients, p-values, and significance (sign.) for SC regression models for algorithm group*

Independent Variable	Coefficients	p-value	Sign.
Intercept	34.1	$1.29{\times}10^{-3}$	**
Trial	1.25	$3.25{\times}10^{-15}$	***
Unsafe Landing	-15.3	$4.45{\times}10^{-4}$	***
Unsuccessful Landings	-8.53	0.0949	
Performance Scores	0.0355	$4.69{\times}10^{-4}$	***
Time per Trial	0.302	$3.86{\times}10^{-6}$	***
Mental Workload	-0.383	$3.87{\times}10^{-9}$	***
Multiple R^2	**0.624**		
Adjusted R^2	**0.616**		

Note: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 4: *Coefficients, p-values, and significance (sign.) for SC regression models for Manual group*

Independent Variable	Coefficients	p-value	Sign.
Intercept	48.1	$2.26{\times}10^{-3}$	**
Trial	1.31	$5.22{\times}10^{-7}$	***
Unsafe Landing	-16.7	$1.99{\times}10^{-3}$	**
Unsuccessful Landings	-3.38	0.676	
Performance Scores	0.0202	0.185	
Time per Trial	-0.110	0.301	
Mental Workload	-0.205	$8.51{\times}10^{-3}$	**
Multiple R^2	**0.265**		
Adjusted R^2	**0.250**		

Note: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 5: *Coefficients, p-values, and significance (sign.) for self-confidence regression models **with assistance independent variable for the algorithm group.***

Independent Variable	Coefficients	p-value	Sign.
Intercept	28.3	$1.29{\times}10^{-3}$	**
Trial	0.661	$3.25{\times}10^{-15}$	***
Unsafe Landing	-12.2	$4.45{\times}10^{-4}$	***
Unsuccessful Landings	-1.70	0.0949	
Performance Scores	0.0560	$4.69{\times}10^{-4}$	***
Time per Trial	0.168	$3.86{\times}10^{-6}$	**
Mental Workload	-0.0270	$3.87{\times}10^{-9}$	***
Assistance On	-21.5	$<2.00{\times}10^{-16}$	***
Multiple R^2	**0.722**		
Adjusted R^2	**0.715**		

Note: *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Part II: Multi-variate Regression Analysis

- Both multiple and adjusted R-squared values are **higher for the algorithm group**. The algorithm group is more likely to consider the given feedback when self-assessing performance.
- R-squared value for algorithm group **increases with the addition of control mode as an independent variable**. The algorithm group self-confidence is impacted significantly by the intelligent automation assistance in addition to the given performance feedback.

CONCLUSIONS

- The independent t-test results show that **participants receiving assistance based on the algorithm achieved similar performance and self-confidence to that of the manual group with less mental workload**.
- Multi-variate regression analysis shows that participants in the algorithm group **exhibit better self-awareness of performance**, and in turn**, better self-efficacy behavior** in the quadrotor landing simulator module.

We can confirm that participants who **received assistance from automation** in the quadrotor simulator **demonstrated more self-efficacy and less mental workload** than those who did not have access to assistance.