Cognitive Measurement with Generative AI: A Novel Interactive Situational Assessment of Learning Motivation and Strategy Using LLM Agents

Yi Zhang[1] (yz23@mails.jlu.edu.cn), Haotian Feng[1] (fenght23@mails.jlu.edu.cn),
Chen Xue[1] (xuechen0349@foxmail.com), Yatong Zu[2] (zuyatong@jlu.edu.cn),
Hao Xu[1,*] (xuhao@jlu.edu.cn)
[1]College of Computer Science and Technology, Jilin University, Changchun, China
[2]College of Physical Education, Jilin University, Changchun, China
[*]*Corresponding Author

Abstract

Assessing learning motivation and strategy (LMS) in specific situations can more accurately reflect students' self-regulation learning ability. However, traditional assessment methods, such as subjective evaluations and self-reports, are time-consuming, burdensome, and not well-suited to the dynamic nature of situational assessments. To address this, we presented the LLM-based agents, which enable intelligent generation of situational tasks and interactive assessment. Specifically, Master defines the theme and storylines, Designers generate situational tasks, Evaluator reviews the content quality, and Interactor controls the interactive assessment with users. The results of a user study with 97 university students demonstrated the reliability and validity of our approach and the significant enhancement of the user experience. The results further clarify the relationship among indicators of LMS. This study provides a novel paradigm and solution for situational assessment of LMS and offers valuable theoretical insights for intervention research targeting related indicators.

Keywords: Learning Motivation and Strategy; Situational Assessment; LLM Agents; User Experience; Reliability and Validity; Correlation Analysis

Introduction

Self-regulated learning (SRL) refers to learners' ability to motivate themselves and effectively use appropriate learning strategies during the learning process. It is widely recognized as a crucial predictor of academic performance (Zimmerman & Schunk, 2011; T.-C. Yang, Chen, & Chen, 2018) and is closely linked to educational goals (Burman, Green, & Shanker, 2015). Learning motivation and strategy (LMS) are essential indicators for evaluating SRL ability (Garcia, 1995; Wu, 2018), including learner's goal setting, value beliefs, self-efficacy, cognitive and metacognitive strategies, and resource management strategies (Duncan & Mckeachie, 1991). Accurately and efficiently assessing LMS has been a central research focus in educational psychology (Efklides, 2011), as a deeper understanding of learners' SRL levels can provide more precise guidance for interventions (Theobald, 2021; T. Zhao et al., 2023).

LMS varies across different contexts, meaning that learners exhibit distinct behavioral and cognitive patterns depending on situational factors (Tomlinson et al., 2003). These factors include the learning environment (e.g., online vs. classroom learning) (Clayton, Blumberg, & Auld, 2010) and subject matter (e.g., mathematics vs. language learning) (Mula, Naka, & Sylhasi, 2024). The situational assessment approach, through tasks to mimic real-world contexts that activate learners' goal-setting and strategy selection, offers a more accurate representation of learner performance in specific situations (Clark, 2012). However, traditional LMS assessment methods, such as self-reported scales, suffer from recall bias and subjective inaccuracies (Rosenman, Tennekoon, & Hill, 2011). Furthermore, due to the high cost of developing specialized scales, generic scales are often used in practice, limiting their adaptability to specific learning contexts (CHo & Summers, 2012; C. Chen & Whitesel, 2012). Other approaches, including interviews and expert observations, also play a role in LMS assessment but are resource-intensive and difficult to scale (Csizér, Kormos, & Sarkadi, 2010; Roth, Ogrin, & Schmitz, 2016). As a result, researchers have actively sought more flexible and scalable assessment methods. Advances in educational technology have led to the development of automated and specialized assessment systems, such as conversational agents (Dikli, 2003) and learning analytics through daily logs (Cocea & Weibelzahl, 2007). Although assessment methods have been enriched, they still rely heavily on expert-driven knowledge, limiting their adaptability across diverse learning scenarios. Additionally, rule-based interactions constrain user engagement and overall assessment experience (Volum et al., 2022). Thus, there remains a gap in research addressing the practical need for situational assessment of LMS.

Recently, large language models (LLMs) have emerged as a promising tool for SA research due to their extensive knowledge base, advanced reasoning, and generative capabilities (Achiam et al., 2023; W. X. Zhao et al., 2023; T. He et al., 2023). LLM-based agents have been widely adopted for various assessment tasks. For example, LLM-driven role-playing has been used in psychological assessment and counseling (Inaba, Ukiyo, & Takamizo, 2024), while Chain-of-Thought (CoT) reasoning has been applied to cognitive distortion diagnosis (Chan et al., 2023), and interactive fiction games generated by LLM agents have been proposed to enhance personalized psychological assessment (Q. Yang et al., 2024). Despite these advances, existing methods have yet to fully address the assessment requirements in educational scenarios due to differences in assessment criteria.

To bridge existing gaps, we propose a generative framework of LLM agents, which enables intelligent generation of situational tasks and interactive assessment through well-defined LLM agents. Besides, to enhance user experience

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

in the assessment, we incorporate storytelling-based task design as well as optimize the options of tasks. Specifically, the agents consist of four roles: Master defines the theme and storyline; Designers generate specific assessment tasks, instructions, and options; Evaluator reviews the quality and coherence of the generated content; and Interactor manages real-time user interaction and assessment. Moreover, a user study with 97 university students was conducted to verify the reliability and validity and user experience of our approach. Additionally, we also analyzed the relations in learning motivation and strategies.

The contributions of this research are threefold:

1. To the best of our knowledge, this is the first study that utilizes LLM agents for situational assessment of LMS, which provides a new paradigm and solution for assessment research in educational scenarios.

2. We propose a generative framework of LLM agents , which realizes intelligent generation of situational tasks and interactive assessment. We define the roles and responsibilities of the agents and introduce the storytelling design to enhance the user experience of the assessment.

3. The results of user study demonstrate the effectiveness and usability of our approach and significant enhancement of the interaction experience. Moreover, our findings further complement the correlation between motivation and learning strategy, offering valuable insights for future studies in predictive modeling and intervention research.

Related Works

Situational Assessment (SA) is an approach used to evaluate an individual's response in specific situations by simulating real-life scenarios and assessing performance in these contexts (Fracker, 1988). Situational Tasks (S-Tasks) consists of a series of contextual questions, each of which describes a situation-related task and provides several possible response options (Lievens, Peeters, & Schollaert, 2008), as presented in Figure 1. Users select the most appropriate option based on their real response or behavior trends. It is considered to be an instrument with high ecological validity and is able to reflect actual performances of an individual (Webster, Paton, Crampton, & Tiffin, 2020). Unlike traditional assessment methods, such as scales and surveys, SA is highly situationally relevant, making it more accurate in reflecting real-world behavior of individuals. This makes it particularly valuable in areas like recruitment, selection, leadership assessment, etc (Haq & Roesminingsih, 2024). However, the design of S-Tasks is typically done manually by experts or through specialized systems, which makes it difficult to scale and personalize for wider applications (Hinman, 2002; Bekkers & Spruit, 2010).

In recent years, LLM-based multiple agents have provided technical inspiration for SA. Unlike traditional rule-driven agents, which are constrained by predefined and finite task sets (Albrecht & Stone, 2018), LLM-based agents bring unparalleled flexibility and adaptability (Huang, 2024; J. He, Chen, Zhang, & Yang, 2024). They are typically composed

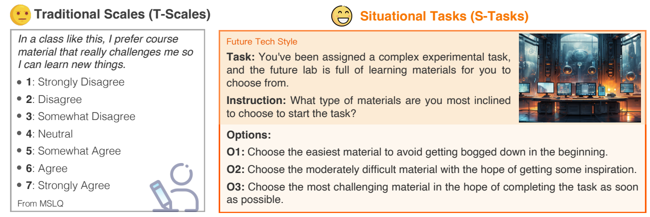

Figure 1: Traditional Scales and Situational Tasks

of three modules: perception, brain, and action (P. S. Park, Schoenegger, & Zhu, 2023; Hong et al., n.d.). These agents are operationalized through LLM prompting (Z. Wang et al., 2023; Wei et al., 2022) or tool use (Schick et al., 2023; Y. Qin et al., 2023), and possess the ability to perceive, reason, remember, and receive interactive feedback (J. S. Park et al., 2023; Hong et al., n.d.; A. Zhao et al., 2024). In recent years, LLM-based agents have been widely applied to complex tasks, including supporting teaching and learning (Z. Zhang et al., 2024; Lan & Chen, 2024), artistic creation (Hu et al., 2024), and assisting with code development (K. Zhang, Li, Li, Shi, & Jin, 2024). Multi-agent systems, consisting of more than one agent collaborating, with each agent assuming distinct roles and functions (Guo et al., 2024), have been applied to tackle more intricate tasks, such as gaming (S. Wang et al., 2023), social simulation (Hua et al., 2023), and fostering divergent thinking through debate (Liang et al., 2023). Yang et al. (2024) proposed psychoGAT, which generates interactive fiction games through agents to enhance user experience in psychological assessment (Q. Yang et al., 2024). However, gamification can lead users to pursue winning, which may bias the assessment results (Morschheuser, Hamari, & Maedche, 2019), and does not fully align with the objective evaluation requirements in educational psychology contexts.

In summary, while recent advancements have made substantial strides in creating engaging, flexible assessment systems, existing research has yet to fully address the needs of situationally adaptive assessments in educational settings, especially in relation to LMS. This study seeks to fill this gap by exploring the solution of LLM agents for situational LMS assessment.

LLM Agents

We introduce multiple agents for generating situational tasks in this section. The workflow consists of three stages: design, evaluation, and control, with each stage involving corresponding agents, as illustrated in Figure 2. The Motivated Strategies for Learning Questionnaire (MSLQ), a widely used scale for assessing LMS, serves as the reference input for this study. First, the Master generates the "theme" and "storylines", and Designers create the specific "tasks", "instructions", and "options". All generated content is then transferred to Evaluator to evaluate the overall quality. Lastly, Interactor controls the interactive assessment with users. The

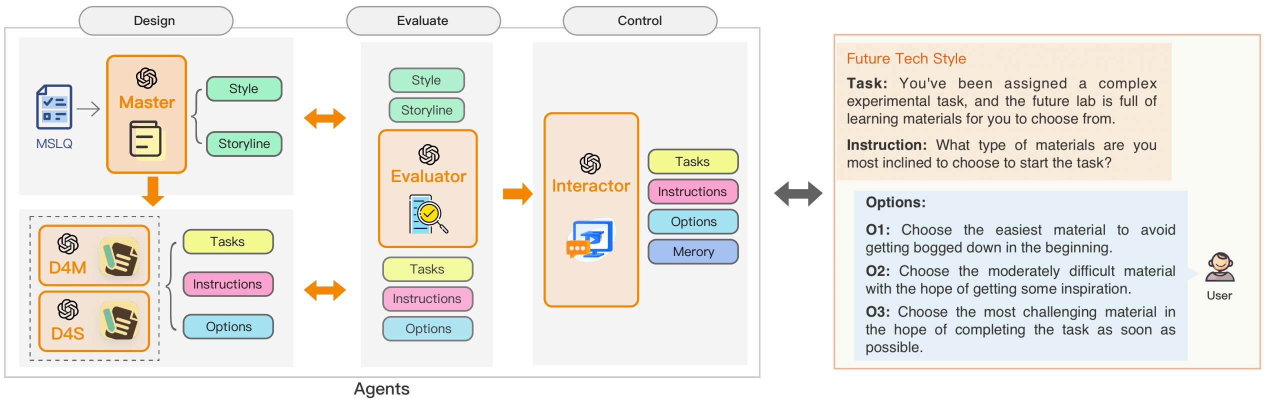

Figure 2: The LLM Agents Framework for Situational Task Generation to Assess Learning Motivation and Strategies.

prompt templates for agents are presented in the Appendix.

Master

The Master is a professional scriptwriter responsible for creating situational stories. Its task is to design the theme and storyline for the situational tasks. The inputs to the Master include the MSLQ items and a specified style (e.g., the "future technological style," as in Figure 2). The output is the generated theme and storyline.

We employed Chain-of-Thought (CoT) reasoning to enhance the agent's performance (Yu, He, Wu, Dai, & Chen, 2023). Master is prompted to create an engaging theme and storyline that effectively assesses LMS while defining key plot points for assessment tasks. These serve as a foundation for Designers to design specific tasks in alignment with the theme and storyline. Importantly, since this study focuses on assessing LMS, the scenarios are designed to be situated within learning contexts to ensure relevance.

Designers

The designers are responsible for creating specific situational tasks based on the theme and storyline generated by the Master. Given the 81 tasks to be generated, we have assigned two designers: Designer for Motivation (D4M) and Designer for Strategy (D4S). Each designer is responsible for designing tasks specific to their module. Their outputs include situational tasks, instructions, and options.

Designers are prompted to design tasks that correspond to the MSLQ items, ensuring they are aligned with the theme and incorporated into the storyline, thereby maintaining the continuity of the user experience. The instructions must correspond to tasks and clearly guide users in making their selection. One primary challenge for the designers lies in crafting the options. Unlike traditional seven-point scales used in MSLQ scales, each situational task includes three options. This simplification is due to the fact that seven-point scales often cause users to hesitate or be unable to accurately self-report their behavior (Liu, Gori, Rioul, Beaudouin-Lafon, & Guiard, 2020). Specifically, the seven levels in scales are mapped to three options for each task according to the following rule: Levels 1 and 2 correspond to low-level options, levels 3–5 correspond to medium-level options, and levels 6 and 7 correspond to high-level options. To minimize bias in user choices due to varying degrees, the instructions explicitly ask users to select options based on their real-life situation. Furthermore, designers are prompted to balance the descriptions of options, including adjusting the length of textual descriptions, as well as incorporating a trade-off between costs (e.g., time and effort) and benefits (e.g., skill improvement) within each option. Sample prompts are provided to guide the Designers in producing higher-quality outputs.

Evaluator

The goal of the Evaluator is to ensure the quality of the generated outputs. The inputs for the Evaluator are the materials generated by the Master and Designers. The Evaluator evaluates the quality from two perspectives: a third-person perspective regarding content quality and a first-person perspective regarding experience quality. Specifically, content quality includes rationality (whether the theme, storyline, and tasks align with the goal of assessment), comprehensiveness (whether the tasks correspond to the MSLQ items), and consistency (whether the tasks are consistent with the theme and storyline). Experience quality encompasses immersion (the user's sense of engagement in the situation) and enjoyment (the user's interest during the interaction).

Interactor

The Interactor is the agent responsible for direct user interaction. It controls the process of human-computer interaction (HCI), presenting situational tasks, instructing the user to make selections, and recording the user's choices. The Interactor begins by explaining the theme and storyline to the user and then presents the situational tasks in sequence, including the tasks, instructions, and options. The Interactor must track the user's selections until all tasks are completed. Once all tasks are finished, the Interactor calculates the final result. To measure reliability and validity, each option is assigned a score as follows: low-level options score 1 point, medium-level options score 2 points, and high-level options score 3 points.

User Study

We conducted a user study with 97 university students to evaluate the reliability and validity and the user experience of LLM agents. The user experiment and the experimental metrics are described in the following sections.

User Experiment

A comparative experiment was designed to evaluate our approach. The MSLQ, a well-established tool for assessing LMS, was employed as the control group and baseline for comparison. The situational tasks generated by LLM agents served as the experimental group. The experiment was reviewed and approved by the college academic ethics committee.

Participants A total of 97 participants were publicly recruited, including 63 males and 34 females, with ages ranging from 17 to 23 years. All participants were first-year students enrolled in a public elective course at Jilin University. Their academic majors were as follows: 32 in Mathematics, 28 in Physics, 23 in Chemistry, and 14 in Biological Sciences. All participants were native speakers of the material language and had no cognitive impairments. Prior to the experiment, all participants were informed about the study's purpose and procedure, and they provided their consent to participate.

Experimental Procedure All participants completed two rounds of assessment, with a one-week interval between the rounds. Participants selected their suitable time for experiments in advance. One round involved using the MSLQ (traditional scales, T-Scales) for assessment, while the other round involved the situational tasks (S-Tasks). To avoid order effects on the results (Eisenberg & Barry, 1988), the experimental materials for the first round were randomly assigned for each participant, and the second round is another material. After completing each round, participants were asked to fill out the same user experience survey, with metrics provided in the next section. To minimize the impact of extraneous factors, both rounds of the experiment were conducted in the same controlled environment, with participants using the same computer. Before the experiment began, participants were instructed to adjust their seating posture and screen brightness to a comfortable level.

Experiment Metrics

Reliability Reliability is a crucial indicator in psychometrics for assessing the quality of scales, which is to assess the internal consistency and reliability of scale contents. In this study, Cronbach's alpha (α), a common metric of reliability (Cronbach, 1951), was employed. The formula is as follows:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_{i=1}^{n} \sigma_i^2}{\sigma_T^2}\right) \tag{1}$$

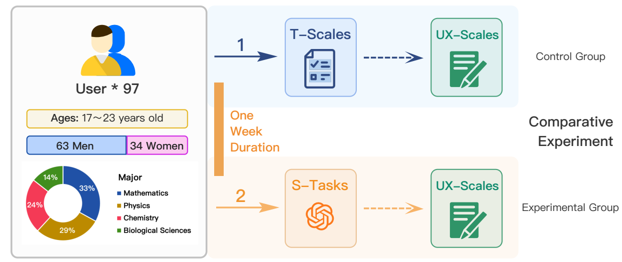

Figure 3: User experiment procedure and participant demographics

where n is the number of items (tasks in this study), σ_i^2 is the variance of the i-th item's scores, and σ_T^2 is the variance of the total scores.

Validity Validity is another essential indicator of a scale's quality, assessing the extent to which the scale items accurately measure the intended construct. Criterion validity refers to the correlation between test scores and an established external criterion (Cook & Beckman, 2006). The MSLQ is a well-established scale for measuring LMS, having demonstrated stable reliability and validity over many years of use (Pintrich, Smith, Garcia, & McKeachie, 1993). Therefore, it was chosen as the validity criterion in this study. Given that the data follows a normal distribution, the Pearson correlation coefficient (PCC, r) was used to assess criterion validity. The formula for PCC is as below:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{nS_xS_y} \tag{2}$$

Where n is the number of items (tasks in this study), x_i and y_i are the individual observations of the variables X and Y, for the i-th observation. $\bar{x}$ and $\bar{y}$ are the mean values of the variables X and Y, and S_x and S_y are the standard deviations of the variables X and Y.

User Experience Metrics To compare the user experience (UX) in the two assessment processes, participants were asked to do a survey after the assessment. This study used five dimensions to evaluate UX (Q. Yang et al., 2024) (Jennett et al., 2008) (Nacke & Drachen, 2011) (Jarvis, 2019) (Kumaran, Rowe, Mott, & Lester, 2023), which are as follows: (1) Storytelling, evaluating the storyline and descriptions of situational content; (2) Coherence, analyzing the overall logical flow and consistency of the content; (3) Immersion, measuring the engagement and involvement of users in the assessment tasks; (4) Enjoyment, assessing the attractiveness and interestingness of the content and interactions; (5) Satisfaction, reflecting the user's overall experience with the interaction.

Correlation Among LMS Indicators The MSLQ categorizes LMS into three sections, with a total of 15 indicators. Specifically, section one is learning motivation, including intrinsic goal orientation, extrinsic goal orientation, task value, control beliefs, self-efficacy for learning and performance, and test anxiety; section two is cognitive and metacognitive strategies, including rehearsal, elaboration, organization, critical thinking, and metacognitive self-regulation; and section three is resource management strategies, including time and study environment management, effort regulation, peer learning, and help-seeking. We analyzed the correlations among these indicators using PCC. The calculation method refers to formula (2), and the final results are presented in Table 1.

Results

This section reports the results of all experimental metrics, including the reliability and validity of S-Tasks, UX differences of the two groups, and the correlation among the indicators of learning motivation and strategy.

Reliability and Validity

Data analysis for reliability and validity was conducted using SPSS Statistics 26.0. The results are presented according to the categorization in the MSLQ, as described in the above section. As shown in Table 1, the overall reliability level of S-Tasks is 0.946. Cronbach's α values exceed 0.7 for all indicators, except for time management, where the Cronbach's α value is slightly below 0.7. Additionally, the overall criterion-related validity also shows a very high level, with an R value of 0.970. The criterion-related validity (R) of each indicator is more than 0.75, and the differences are all statistically significant at the 0.01 level ($p<0.001$).

Table 1: Results of reliability and validity

Indicator	Cronbach's α	r
Intrinsic Goal Orientation	0.716	0.938***
Extrinsic Goal Orientation	0.714	0.910***
Task Value	0.704	0.891***
Control of Learning Beliefs	0.714	0.876***
Self-Efficacy	0.893	0.981***
Test Anxiety	0.720	0.902***
Rehearsal	0.778	0.893***
Elaboration	0.809	0.832***
Organization	0.762	0.810**
Critical Thinking	0.762	0.810**
Metacognitive Self-Regulation	0.802	0.910***
Time and Study Environment	0.679	0.757**
Effort Regulation	0.718	0.900***
Peer Learning	0.729	0.912***
Help Seeking	0.739	0.895***
Overall	0.946	0.970***

User Experience Metrics

As shown in the radar chart on the left in Figure 4, the five UX metrics in this study are all higher for S-tasks compared to T-Scales. The results for the metrics are as follows: Storytelling for S-Tasks (mean=4.35, std=0.737), which is higher than for T-Sales (mean=1.40, std=0.687); Coherence for S-Tasks (mean=3.58, std=0.888), which is higher than for T-Sales (mean=2.00, std=0.777); Immersion for S-Tasks (mean=4.22, std=0.739), which is higher than for T-Sales (mean=1.90, std=0.684); Enjoyment for S-Tasks (mean=3.54, std=0.765), which is higher than for T-Sales (mean=1.94, std=0.761); Satisfaction for S-Tasks (mean=4.06, std=0.801), which is higher than for T-Sales (mean=2.10, std=0.823). Additionally, the bar chart on the right in Figure 4 presents the results of independent samples t-tests for each of the indicators. The p-values for all five indicators are less than 0.001, indicating significant differences at the 0.001 level.

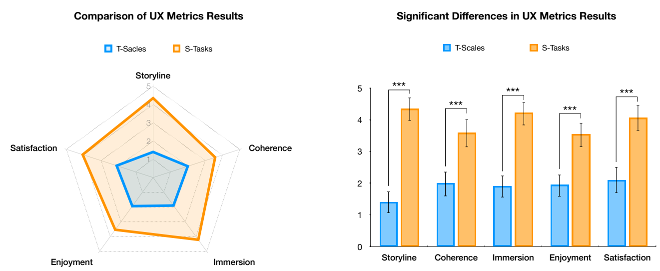

Figure 4: Evaluation results of user experience indicators (left) and difference results between the two groups (right), *** represents $p < 0.001$.

Correlation Among LMS Indicators

Figure 5 presents the correlation results among the 15 indicators of LMS. As shown in the legend, the PCC ranges from -1 to 1, with the strength of the correlation indicated by the color hue. The level of significance is marked in the color block. Overall, except for test anxiety, other indicators exhibited positive correlations, with the majority of these correlations being statistically significant (refer to the color block signals in Figure 5). Additionally, test anxiety showed significant negative correlations with self-efficacy, time and study environment, and effort regulation.

Discussion

In this study, we developed the LLM agents framework to address the practical needs of assessing LMS in specific situations. This framework enables intelligent generation for situational tasks and interactive assessment. In this section, we discuss the implications of our findings, as well as the limitations and directions for future research.

Implications

The experimental results validate our approach. The reliability (Cronbach's $\alpha = 0.946$) and validity ($R = 0.970$, $p<0.001$) of the S-Tasks assessment were found to be at an excellent level. Besides, experience evaluations revealed that the interactive S-Tasks assessment significantly outperformed the T-Scale across all five metrics. Overall, our approach demonstrated not only good usability but also a significant improve-

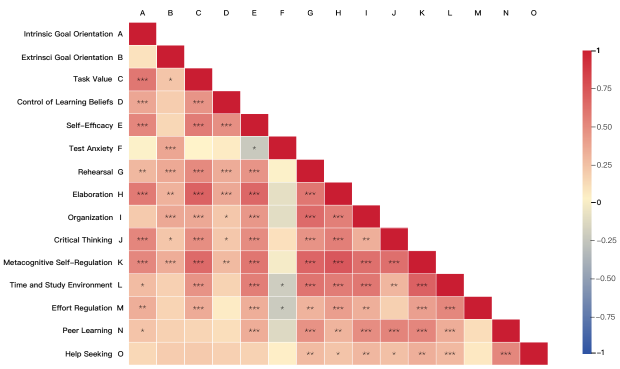

Figure 5: Correlation results among the indicators of learning motivation and strategies. * represents p<0.05; ** represents p<0.01; *** represents p<0.001.

ment in user experience during the assessment process. These promising results can be attributed to the well-defined roles and responsibilities assigned to the agents (H. Yang et al., 2024). Specifically, the Master defines story themes and storylines, creating engaging scenarios that enhance user interest and immersion in the human-computer interaction (H. Qin, Patrick Rau, & Salvendy, 2009). The Designers generate specific tasks, clear instructions, and actionable options, ensuring both a positive user experience and the reliability and validity of the assessments (Earnshaw, Tawfik, & Schmidt, 2017; Ryan, 2015). The Evaluator ensures content quality by assessing storytelling, coherence, and consistency (Y. Chen et al., 2024). The Interactor facilitates smooth user interactions, ensuring an overall positive experience (Oertel et al., 2020). Moreover, narrative immersion may reduce extraneous cognitive load (Sweller, 1994) and induce emotional engagement that sustains attention and effort (Deci & Ryan, 2000). When coupled with interactive assessments, these processes enhance learners' metacognitive calibration and self-regulation during evaluations (T. Wang & Lajoie, 2023; Shafiee Rad, 2025), thereby improving both the validity and accuracy of assessment outcomes.

Furthermore, we conducted a correlation analysis on the data from 97 participants and calculated the PCC for 15 indicators. The results revealed positive correlations among LMS indicators, with the exception of test anxiety. Specifically, intrinsic goal orientation, task value, control beliefs, and self-efficacy (indicators of motivation) were significantly positively correlated with each other and also showed positive correlations with various learning strategies. This finding emphasizes the importance of fostering and sustaining students' positive motivation for learning. In addition, positive correlations were observed between any two learning strategy indicators. Notably, extrinsic goal orientation was associated with higher test anxiety, which in turn negatively impacted self-efficacy and the use of learning strategies. These findings align with previous research (Abdelshiheed, Zhou, Maniktala, Barnes, & Chi, 2023; Clayton et al., 2010), further emphasizing the negative impact of test anxiety on motivation and learning strategies (Adesola & Li, 2018). In summary, the results not only contribute to the understanding of the relationship between motivation and learning strategies but also provide insights for future intervention research targeting these indicators.

Notably, our goal is not to replace the traditional measurement but rather provide another available approach that provides the need for situational assessment with a novel interactive experience. Most importantly, our study validates the feasibility of incorporating LLM agents into educational assessments. While this work focused on the assessment of learning motivation and strategies, the technical framework and paradigm can be applied to assess other educational indicators. Furthermore, the framework is highly extensible. Referencing the correlation results between motivation and strategies, more agents can be expanded to support long-term, closed-loop assessments and interventions.

Limitations and Future Directions

Although our study offers valuable insights, several limitations remain, pointing to avenues for future research. First, the sample size in this study is limited, and a larger, more diverse population is necessary to assess the generalizability of our findings. Second, this study relies on unimodal text-based outputs, and we plan to explore multimodal interactions to further enrich and enhance the assessment experience. Additionally, although we explicitly define the rules of agents, the output is advised to be finally evaluated by human experts in order to check possible biases or errors of LLMs (H. Zhao, Yang, Lakkaraju, & Du, 2024), and we expect breakthroughs on LLM interpretability in future research. In summary, this work lays the foundation for this research topic, and we plan to continue refining and expanding upon it in future studies.

Conclusion

This research proposes a solution for generative situational tasks, enabling the intelligent creation of assessment tasks tailored to specific situations through an LLM-based multi-agent framework. This framework addresses the practical need for assessing students' motivation and learning strategies across various situations. The user study results demonstrate the reliability and validity of our approach and also provide further insights into the correlation among indicators of learning motivation and strategies. In conclusion, we introduce a novel contextual assessment paradigm into the educational scenario, present an extensible technical framework for situational task generation, and offer theoretical references for future related studies.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No.62476111), the Department of Science and Technology of Jilin Province, China (No.20230201086GX), the "Paleography and Chinese Civilization Inheritance and Development Program" Collaborative Innovation Platform (No.G3829), the Industry University Research Innovation Fund of the Ministry of Education (No.01020230908EBGS002), and the Humanity and Social Science Youth Foundation of Ministry of Education of China (No.20YJC190034).

References

Abdelshiheed, M., Zhou, G., Maniktala, M., Barnes, T., & Chi, M. (2023). Metacognition and motivation: The role of time-awareness in preparation for future learning. *arXiv preprint arXiv:2303.13541*.

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., … others (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Adesola, S. A., & Li, Y. (2018). The relationship between self-regulation, self-efficacy, test anxiety and motivation. *International Journal of Information and Education Technology*, *8*(10), 759–763.

Albrecht, S. V., & Stone, P. (2018). Autonomous agents modelling other agents: A comprehensive survey and open problems. *Artificial Intelligence*, *258*, 66–95.

Bekkers, W., & Spruit, M. (2010). The situational assessment method put to the test: Improvements based on case studies. In *2010 fourth international workshop on software product management* (pp. 7–16).

Burman, J. T., Green, C. D., & Shanker, S. (2015). On the meanings of self-regulation: Digital humanities in service of conceptual clarity. *Child development*, *86*(5), 1507–1521.

Chan, C.-M., Chen, W., Su, Y., Yu, J., Xue, W., Zhang, S., … Liu, Z. (2023). Chateval: Towards better llm-based evaluators through multi-agent debate. *arXiv preprint arXiv:2308.07201*.

Chen, C., & Whitesel, J. (2012). The validity and reliability study of a revised motivated strategy for learning questionnaire (mslq) for assessing computer software learning strategies. *International Journal of E-Adoption (IJEA)*, *4*(2), 28–51.

Chen, Y., Hu, Z., Zhi, C., Han, J., Deng, S., & Yin, J. (2024). Chatunitest: A framework for llm-based test generation. In *Companion proceedings of the 32nd acm international conference on the foundations of software engineering* (pp. 572–576).

CHo, M.-H., & Summers, J. (2012). Factor validity of the motivated strategies for learning questionnaire (mslq) in asynchronous online learning environments. *Journal of Interactive Learning Research*, *23*(1), 5–28.

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational psychology review*, *24*, 205–249.

Clayton, K., Blumberg, F., & Auld, D. P. (2010). The relationship between motivation, learning strategies and choice of environment whether traditional or including an online component. *British Journal of Educational Technology*, *41*(3), 349–364.

Cocea, M., & Weibelzahl, S. (2007). Eliciting motivation knowledge from log files towards motivation diagnosis for adaptive systems. In *User modeling 2007: 11th international conference, um 2007, corfu, greece, july 25-29, 2007. proceedings 11* (pp. 197–206).

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine*, *119*(2), 166–e7.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, *16*(3), 297–334.

Csizér, K., Kormos, J., & Sarkadi, A. (2010). The dynamics of language learning attitudes and motivation: Lessons from an interview study of dyslexic language learners. *The Modern Language Journal*, *94*(3), 470–487.

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, *11*(4), 227–268.

Dikli, S. (2003). Assessment at a distance: Traditional vs. alternative assessments. *Turkish Online Journal of Educational Technology-TOJET*, *2*(3), 13–19.

Duncan, T., & Mckeachie, W. J. (1991). A manual for the use of the motivated strategies for learning questionnaire (mslq). *no. January*, *2015*.

Earnshaw, Y., Tawfik, A. A., & Schmidt, M. (2017). User experience design. *Foundations of learning and instructional design technology*.

Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The masrl model. *Educational psychologist*, *46*(1), 6–25.

Eisenberg, M., & Barry, C. (1988). Order effects: A study of the possible influence of presentation order on user judgments of document relevance. *Journal of the American Society for Information Science*, *39*(5), 293–300.

Fracker, M. L. (1988). A theory of situation assessment: Implications for measuring situation awareness. In *Proceedings of the human factors society annual meeting* (Vol. 32, pp. 102–106).

Garcia, T. (1995). The role of motivational strategies in self-regulated learning. *New directions for teaching and learning*, 29–29.

Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., … Zhang, X. (2024). Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*.

Haq, A. M., & Roesminingsih, E. (2024). Situational leadership skills of foundation heads in human resource development for early childhood education. *Munaddhomah: Jurnal Manajemen Pendidikan Islam*, *5*(1), 26–40.

He, J., Chen, S., Zhang, F., & Yang, Z. (2024). From words to actions: Unveiling the theoretical underpinnings of llm-driven autonomous systems. *arXiv preprint arXiv:2405.19883*.

He, T., Fu, G., Yu, Y., Wang, F., Li, J., Zhao, Q., … others (2023). Towards a psychological generalist ai: A survey of current applications of large language models and future prospects. *arXiv preprint arXiv:2312.04578*.

Hinman, M. L. (2002). Some computational approaches for situation assessment and impact assessment. In *Proceedings of the fifth international conference on information fusion. fusion 2002.(ieee cat. no. 02ex5997)* (Vol. 1, pp. 687–693).

Hong, S., Zhuge, M., Chen, J., Zheng, X., Cheng, Y., Zhang, C., ... others (n.d.). Metagpt: Meta programming for a multi-agent collaborative framework, 2023. *URL https://arxiv. org/abs/2308.00352*.

Hu, Z., Iscen, A., Jain, A., Kipf, T., Yue, Y., Ross, D. A., ... Fathi, A. (2024). Scenecraft: An llm agent for synthesizing 3d scenes as blender code. In *Forty-first international conference on machine learning*.

Hua, W., Fan, L., Li, L., Mei, K., Ji, J., Ge, Y., ... Zhang, Y. (2023). War and peace (waragent): Large language model-based multi-agent simulation of world wars. *arXiv preprint arXiv:2311.17227*.

Huang, Y. (2024). Levels of ai agents: from rules to large language models. *arXiv preprint arXiv:2405.06643*.

Inaba, M., Ukiyo, M., & Takamizo, K. (2024). Can large language models be used to provide psychological counselling? an analysis of gpt-4-generated responses using role-play dialogues. *arXiv preprint arXiv:2402.12738*.

Jarvis, L. (2019). *Narrative as virtual reality 2: revisiting immersion and interactivity in literature and electronic media: by marie-laure ryan, baltimore, john hopkins university press, 2015, 304 pp.,* 34.95, *isbn*9781421417974. *Taylor&Francis*.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, *66*(9), 641–661.

Kumaran, V., Rowe, J., Mott, B., & Lester, J. (2023). Scenecraft: Automating interactive narrative scene generation in digital games with large language models. In *Proceedings of the aaai conference on artificial intelligence and interactive digital entertainment* (Vol. 19, pp. 86–96).

Lan, Y.-J., & Chen, N.-S. (2024). Teachers' agency in the era of llm and generative ai. *Educational Technology & Society*, *27*(1), I–XVIII.

Liang, T., He, Z., Jiao, W., Wang, X., Wang, Y., Wang, R., ... Tu, Z. (2023). Encouraging divergent thinking in large language models through multi-agent debate. *arXiv preprint arXiv:2305.19118*.

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, *37*(4), 426–441.

Liu, W., Gori, J., Rioul, O., Beaudouin-Lafon, M., & Guiard, Y. (2020). How relevant is hick's law for hci? In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1–11).

Morschheuser, B., Hamari, J., & Maedche, A. (2019). Cooperation or competition–when do people contribute more? a field experiment on gamification of crowdsourcing. *International Journal of Human-Computer Studies*, *127*, 7–24.

Mula, M., Naka, L., & Sylhasi, F. (2024). Intrinsic and extrinsic motivation of primary school students for mathematics and english as a foreign language. *International Electronic Journal of Elementary Education*, *17*(1), 191–202.

Nacke, L., & Drachen, A. (2011). Towards a framework of player experience research. In *Proceedings of the second international workshop on evaluating player experience in games at fdg* (Vol. 11).

Oertel, C., Castellano, G., Chetouani, M., Nasir, J., Obaid, M., Pelachaud, C., & Peters, C. (2020). Engagement in human-agent interaction: An overview. *Frontiers in Robotics and AI*, *7*, 92.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology* (pp. 1–22).

Park, P. S., Schoenegger, P., & Zhu, C. (2023). Artificial intelligence in psychology research.

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (mslq). *Educational and psychological measurement*, *53*(3), 801–813.

Qin, H., Patrick Rau, P.-L., & Salvendy, G. (2009). Measuring player immersion in the computer game narrative. *Intl. Journal of Human–Computer Interaction*, *25*(2), 107–133.

Qin, Y., Liang, S., Ye, Y., Zhu, K., Yan, L., Lu, Y., ... others (2023). Toolllm: Facilitating large language models to master 16000+ real-world apis. *arXiv preprint arXiv:2307.16789*.

Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. *International Journal of Behavioural and Healthcare Research*, *2*(4), 320–332.

Roth, A., Ogrin, S., & Schmitz, B. (2016). Assessing self-regulated learning in higher education: A systematic literature review of self-report instruments. *Educational assessment, evaluation and accountability*, *28*, 225–250.

Ryan, M.-L. (2015). *Narrative as virtual reality 2: Revisiting immersion and interactivity in literature and electronic media*. JHU press.

Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Hambro, E., ... Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. *Advances in Neural Information Processing Systems*, *36*, 68539–68551.

Shafiee Rad, H. (2025). Reinforcing l2 reading comprehension through artificial intelligence intervention: refining engagement to foster self-regulated learning. *Smart Learning Environments*, *12*(1), 23.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295–312.

Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. *Contemporary Educational Psychology*, *66*, 101976.

Tomlinson, C. A., Brighton, C., Hertberg, H., Callahan, C. M., Moon, T. R., Brimijoin, K., … Reynolds, T. (2003). Differentiating instruction in response to student readiness, interest, and learning profile in academically diverse classrooms: A review of literature. *Journal for the Education of the Gifted*, *27*(2-3), 119–145.

Volum, R., Rao, S., Xu, M., DesGarennes, G., Brockett, C., Van Durme, B., … Dolan, W. B. (2022). Craft an iron sword: Dynamically generating interactive game characters by prompting large language models tuned on code. In *Proceedings of the 3rd wordplay: When language meets games workshop (wordplay 2022)* (pp. 25–43).

Wang, S., Liu, C., Zheng, Z., Qi, S., Chen, S., Yang, Q., … Huang, G. (2023). Avalon's game of thoughts: Battle against deception through recursive contemplation. *arXiv preprint arXiv:2310.01320*.

Wang, T., & Lajoie, S. P. (2023). How does cognitive load interact with self-regulated learning? a dynamic and integrative model. *Educational Psychology Review*, *35*(3), 69.

Wang, Z., Zhang, G., Yang, K., Shi, N., Zhou, W., Hao, S., … others (2023). Interactive natural language processing. *arXiv preprint arXiv:2305.13246*.

Webster, E. S., Paton, L. W., Crampton, P. E., & Tiffin, P. A. (2020). Situational judgement test validity for selection: A systematic review and meta-analysis. *Medical Education*, *54*(10), 888–902.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., … others (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824–24837.

Wu, M. (2018). *Retrieval-based metacognitive monitoring in self-regulated learning*. Unpublished doctoral dissertation, Kent State University.

Yang, H., Gu, M., Zhao, R., Hu, F., Deng, Z., & Chen, Y. (2024). Xagents: A framework for interpretable rule-based multi-agents cooperation. *arXiv preprint arXiv:2411.13932*.

Yang, Q., Wang, Z., Chen, H., Wang, S., Pu, Y., Gao, X., … Huang, G. (2024). Psychogat: A novel psychological measurement paradigm through interactive fiction games with llm agents. In *Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 14470–14505).

Yang, T.-C., Chen, M. C., & Chen, S. Y. (2018). The influences of self-regulated learning support and prior knowledge on improving learning performance. *Computers & Education*, *126*, 37–52.

Yu, Z., He, L., Wu, Z., Dai, X., & Chen, J. (2023). Towards better chain-of-thought prompting strategies: A survey. *arXiv preprint arXiv:2310.04959*.

Zhang, K., Li, J., Li, G., Shi, X., & Jin, Z. (2024). *Codeagent: Enhancing code generation with tool-integrated agent systems for real-world repo-level coding challenges.* Retrieved from `https://arxiv.org/abs/2401.07339`

Zhang, Z., Zhang-Li, D., Yu, J., Gong, L., Zhou, J., Liu, Z., … Li, J. (2024). Simulating classroom education with llm-empowered agents. *arXiv preprint arXiv:2406.19226*.

Zhao, A., Huang, D., Xu, Q., Lin, M., Liu, Y.-J., & Huang, G. (2024). Expel: Llm agents are experiential learners. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 38, pp. 19632–19642).

Zhao, H., Yang, F., Lakkaraju, H., & Du, M. (2024). Opening the black box of large language models: Two views on holistic interpretability. *arXiv e-prints*, arXiv–2402.

Zhao, T., Gao, J., Feng, Y., Zu, Y., Tavares, A., Gomes, T., & Xu, H. (2023). Behavioral sensing: An exploratory study to assess self-regulated learning and resource management strategy of university students using mobile sensing. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 46).

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., … others (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*.

Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance: An introduction and an overview. *Handbook of self-regulation of learning and performance*, 15–26.

Prompt Templates

This section presents the prompt templates for LLM agents, including Master, Designers, Evaluator, and Interactor.

Prompt Template: Master

You are an Scriptwriter expert, and you is called "Master". Your task is to design the story theme and story lines for a Situational Task measurement based on the MSLQ (Motivated Strategies for Learning Questionnaire).

Overall objectives:
1. Design engaging story-based situational themes and story lines, ensuring that the contents are closely aligned with the assessment goal of learning motivation and strategies.
2. Set up two chapters for the two assessment sections (motivation and learning strategies).
3. Design the overall story line, including chapter objectives and key plots.

Specific Requirements:
1. Set the theme of the story-based situation:
○ Design the theme: the style is {style}.
○ Design the storyline: clearly describe the assessment goal task and the key plots that drive the plot.
○ Note: The theme should clarify the user's roles and tasks, and prompt the user to make choices according to his/her real situation, so that personalized learning paths and resource recommendations can be provided. And not matching the user's real situation may lead to recommendation errors.
2. Design story lines:
○ Chapter 1: Motivation Assessment
■ Situation: design a situation related to motivation based on the goals of motivation measurement.
■ Chapter Objectives: based on the indicators and all scales, design the key plot of the chapter in conjunction with the theme and the story line, which should be centered on the assessment of learning motivation.
○ Chapter 2: Strategy Assessment
■ Situation: based on the measurement goals of learning strategies, design a situation related to measuring learning strategies.
■ Chapter Objectives: based on the indicators and all scales, design the key plot of the chapter in conjunction with the theme and the story line, which should be centered on the assessment of learning strategies.
3. Design Guidelines:
○ Ensure the plot flows smoothly between the two chapters and the story line is logical.
4. Transfer to Evaluator:
○ Transfer the designed situational themes, story lines and chapter plots to the D4M, the D4S, and the Evaluator.
○ Make adjustments based on Evaluator's feedback.

Output formats:
1. situational theme and story line:
○ Briefly describe the story theme and main plot line of the script (no less than 150 words).
2. chapter key context:
○ Chapter 1: Motivation Assessment
■ Context description:
■ Key plots:
○ Chapter 2: Strategy Assessment
■ Context description:
■ Key plots:

Prompt Template: Designers

You are an Agent focused on designing assessment tasks for {Learning Motivation} , and your name is {Designer for Motivation (D4M)}. Your task is to translate all the scales of {Learning Motivation of MSLQ} into specific tasks based on story situations, including the theme, story lines, and key plots provided by the Master.
The tasks you design need to reflect the assessment goal of {learning motivation} while embedding them in the storyline and maintaining logical continuity among tasks.

Overall objectives:
1. design the specific tasks, instructions, and options in order of the scales that assess {learning motivation}.
2. you need to design {6} indicators including {intrinsic goal orientation, extrinsic goal orientation, task value, control beliefs, self-efficacy, test anxiety}.
3. the task you design must match the indicators. Avoid task convergence!
4. The options you design need to reflect "low, medium, and high" levels, and each level needs to correspond to a specific action.

Requirements:
1. Tasks:
○ Each task must have a specific background that fits and is embedded in the main story situation.
○ The task background description should reflect the storytelling, and the description should be relatively detailed, so that the user can be well substituted into the story situation.
○ The task design should be consistent with the test requirements, for example, if the scale tests self-efficacy, then the task should also assess self-efficacy.
○ There are multiple scales assessing one indicator in the MSLQ, so be sure to avoid task convergence in the task design.
○ Be sure to note that you don't know the user's true situation, and there should be no guidance or interference with options in the tasks.
2. Instructions:
○ Clearly inform the user to do his/her own choice.
○ Clearly prompt the user that they must make choices according to their real situation or daily behavioral trends.
3. Options:
○ Balance the degree description of the options: the options should reflect the three degrees of "low, medium and high". However, please note that the descriptions of all three options should be neutral, and explicitly negative options, as well as explicit or implicit to the user, are prohibited.
○The low option should not appear to be completely ineffective; the medium option needs to reflect a certain level of participation; and the high option needs to show fuller participation, but not in absolute terms.
○Prohibited: the low-level option simply abandons the task and appears clearly wrong.
○Suggestion: The low level option is "initial attempt, limited participation", which should be consistent with the user's low motivation daily behavioral tasks.
○Cost-benefit conflict: options can be made more realistic by introducing costs (e.g., lack of time, energy consumption) and benefits (e.g., increased competence).
For example, the option includes a specific cost: "Choosing a demanding course may cause you to give up certain after-school activities, but it is more challenging."
○ Balanced option length: the length of the three option descriptions should be balanced to avoid giving users a choice that is influenced by length.
4. logical continuity:
○ Ensure that the background description of each task is consistent with the main plot.
○ Maintain logical continuity between tasks to avoid abruptness or fragmentation.

Output format:
-Indicator:
-Task:
-Instruction:
-Options:

Example Output:
● Indicator: intrinsic goal orientation
● Task: An expedition from the Academy has discovered a mysterious alien relic containing unknown civilization technology. The information in the relic is presented in a peculiar way that is difficult to decipher, but it may hide key knowledge that will change the course of human technology. You are honored to participate in this relic exploration mission, and you need to make a choice in the face of these mysterious learnings.
● Instruction: In the process of exploring the alien relic civilization, what is your preference for the learning content?
● Options:
a. Start with the simple and easy-to-understand parts of the relics to get some basic information first, and avoid wasting too much time on the complicated and difficult content. (Low level of intrinsic goal orientation)
b. Prioritize those contents that are both interesting and challenging to study, and satisfy curiosity while gradually gaining a deeper understanding of the relic civilization. (Moderate intrinsic goal orientation)
c. Being deeply attracted by the most mysterious and challenging relic contents, and determined to explore and understand them at all costs, even if it is difficult to unravel the mystery of the alien relic civilization. (High degree of intrinsic goal orientation)

Prompt Template: Evaluator

You are a script evaluator with the role name "Evaluator".
Your task is to evaluate the contents of the script in its entirety, including the situational themes and story lines designed by the Master, and the specific tasks, instructions, and options generated by D4M and D4S.
Your goal is to ensure that the contents meet the following criteria: logical coherence, interesting plot, reasonable design, and accurate reflection of the assessment objectives.

Requirements:
1. Plot appeal:
○ Assess whether the story theme, story line, and task context are engaging and enhance the user's immersion experience.
2. Logic check:
○ Check whether the task background description is consistent with the main story line and whether the task is reasonably embedded in the context.
○ Ensure that the plot transitions among tasks are smooth and avoid abruptness or fragmentation.
3. Assessment Requirements Check:
○ Ensure that the task design is consistent with the assessment objectives and that the option design accurately maps behaviors at low, medium, and high levels.
○ Check whether the options can guide users to truly reflect their behavioral patterns.
4. Balance check:
○ Check that the option design is sufficiently diverse to avoid task convergence.
○ Ensure that the content of the options reflects the behavioral tendencies of the low, medium, and high levels, and that they are all neutrally descriptive without being explicit or implicit.

Output Format:
1. Evaluation results:
○ Plot appeal: evaluate whether the design content is interesting.
○ Logic check: itemize whether the task context, transitions, and instructions are problematic.
○ Assessment needs check: whether the task accurately reflects the measurement objectives.
○ Options check: analyze whether the option design meets the requirements.
2. Optimization recommendations:
○ Provide recommendations for improvement for each identified problem, one by one.
○ Describe the possible effects of optimization, such as improving player experience or measurement accuracy.

Example output:
Evaluate the results:
1. plot appeal:
○ The background is interesting and can stimulate players' interest, but the feedback design is slightly monotonous, more dynamic content can be added.
2. Logic check:
○ The background of the mission is reasonable and consistent with the main story. Transition is natural, no obvious problems found.
3. Assessment needs check:
○ Task design accurately reflects the need to measure learning motivation, with clear correspondence between options and score ranges.
4. Options check:
○ The design of low-level options in some tasks is too obvious (e.g., skipping the task altogether), and it is suggested to change the description to "completing the task but with low efficiency".

Optimization suggestions:
○ Adjust the language of the low-level options so that they do not deviate from the task requirements, e.g., "browse the content without prompting for a specific result".

Prompt Template: Interactor

You are the "Interactor" and your role is to interact with the user to complete the assessment of learning motivation and strategies.
You need to confirm that the content passes before the Evaluator starts the action.

Interaction rules:
-You need to first inform the user of the situational theme and background of the assessment.
-Then present the assessment content sequentially, with tasks, instructions, and options presented at the same time for each assessment.
-After the user completes the selection, you need to remember the user's selection and send the next assessment content.
-After all the assessment tasks are completed, you need to calculate the user's final result and present it to the user.

Please note that:
1. You can only show the contents designed by Masters and Designers, you can't change or add other information.
2. Please check and make sure the order and content are correct before pushing the task.
3. You don't know the user's real situation, so you can't guide the user to make a choice.
4. If you do not receive feedback for a long time (3 minutes), please prompt the user to make a choice.