Scaffolding Language Learning via Multi-modal Tutoring Systems with Pedagogical Instructions

Zhengyuan Liu[*❖], Stella Xin Yin[*✦], Carolyn Lee[♠], Nancy F. Chen[❖]
[✦]Nanyang Technological University, Singapore [♠]Stanford University
[❖]Institute for Infocomm Research (I^{2}R), A*STAR, Singapore

***Abstract*—Intelligent tutoring systems (ITSs) that imitate human tutors and aim to provide immediate and customized instructions or feedback to learners have shown their effectiveness in education. With the emergence of generative artificial intelligence, large language models (LLMs) further entitle the systems to complex and coherent conversational interactions. These systems would be of great help in language education as it involves developing skills in communication, which, however, drew relatively less attention. Additionally, due to the complicated cognitive development at younger ages, more endeavors are needed for practical uses. Scaffolding refers to a teaching technique where teachers provide support and guidance to students for learning and developing new concepts or skills. It is an effective way to support diverse learning needs, goals, processes, and outcomes. In this work, we investigate how pedagogical instructions facilitate the scaffolding in ITSs, by conducting a case study on guiding children to describe images for language learning. We construct different types of scaffolding tutoring systems grounded in four fundamental learning theories: knowledge construction, inquiry-based learning, dialogic teaching, and zone of proximal development. For qualitative and quantitative analyses, we build and refine a seven-dimension rubric to evaluate the scaffolding process. In our experiment on GPT-4V, we observe that LLMs demonstrate strong potential to follow pedagogical instructions and achieve self-paced learning in different student groups. Moreover, we extend our evaluation framework from a manual to an automated approach, paving the way to benchmark various conversational tutoring systems.**

***Index Terms*—Intelligent Tutoring Systems, Scaffolding, Multimodal Language Models**

I. Introduction

Intelligent Tutoring Systems (ITSs) are adaptive instructional systems equipped with Artificial Intelligence (AI) and Natural Language Processing (NLP) technologies and integrated educational methodologies [1]. These systems offer personalized learning content, instant feedback, and interactive learning experience to learners. A significant feature of ITSs is their ability to tailor instructional activities and strategies to align with the learners' different characteristics, experiences, and learning needs [2]. Numerous studies have highlighted the effectiveness and broad applicability of ITSs across various educational fields [1], [3]. On the other hand, with the emergence of generative artificial intelligence [4], [5], large language models (LLMs) further empower interactive ITSs with exceptional capabilities on conversational interactions [6], [7], and show great potential to support students learning outside of classrooms in various disciplines, and they can make online education personalized and more accessible [8].

Fig. 1. A dialogue example of interactive language learning via an image description tutoring system. The student is asked to describe the picture.

For the applications of ITSs, the majority of the studies were conducted within the field of computer science and mathematics education and primarily targeted for university students [1]. Compared to those fields which more focus on abstract concepts, learning a language involves developing skills in communication, including speaking, listening, reading, and writing, resulting in a higher interaction demand for ITSs. Given the diverse learning needs, the complexity of open-ended questions, and cognitive development at younger ages, it's more challenging for ITSs to be applied effectively in children's language education.

"It is only when scaffolding is needed that learning will actually take place (Gibbons, 2002)"

To improve tutoring systems in language learning, a crucial aspect is providing effective scaffolding for learners. Scaffolding refers to an instructional technique in which teachers provide temporary and dynamic support and guidance to students as they learn and develop new concepts or skills [9]. Over

* Equal Contribution. This research is supported by the Agency for Science, Technology and Research (AI4EDU Programme), and the National Research Foundation, Singapore under its AISG Programme (AISG2-GC-2022-005).

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00223

recent decades, scaffolding has been promoted as an effective way to support diverse learning needs, goals, processes, and outcomes. Its application is particularly effective in children's language learning [10]. Previous studies have shown that learners are more likely to succeed in language learning when their teachers provide pedagogical support, facilitating a higher level of skill and understanding [10], [11].

In practice, teachers provide scaffolding by clarifying, questioning, and presenting models for the learners. They guide students through hints and suggestions, encouraging them to connect new information with their personal experiences and prior knowledge [10]. Since scaffolding is a dynamic intervention finely tuned to the learner's ongoing progress, the support given by the teacher during scaffolding strongly depends on the patterns of teacher-student interactions [12]. Therefore, it would be of great help to improve the performance of ITSs by promoting the effectiveness of interactions.

In this work, we investigate how pedagogical instructions facilitate the scaffolding in LLM-based ITSs, and conduct a case study on guiding children to describe images for language learning. Following task-specific prompts, multi-modal LLMs can effectively organize the conversation by asking proper questions, giving constructive suggestions, and providing informative hints. We build different types of conversational ITS grounded in four fundamental learning theories: knowledge construction, inquiry-based learning, dialogic teaching, and zone of proximal development, and compare them in simulated teaching sessions of a set of images and two student groups. To evaluate the scaffolding process of conversational interactions, we build and refine a seven-dimension rubric, which is employed in both qualitative and quantitative evaluations through feedback, hints, instruction, explaining, modeling, questioning, and social-emotional support. In our experiments on GPT-4V, we observe that LLMs can demonstrate strong potential to follow pedagogical instructions and achieve self-paced learning in different student groups. Furthermore, we transform our evaluation framework from manual to automated by leveraging the in-context learning capability of LLMs, and this paves the way to benchmark various tutoring systems.

II. Related Work

A. Intelligent Tutoring Systems

ITSs aim to provide personalized and effective instructional support to students, and they have gained increasing importance due to the growing demand for adaptive and accessible education, especially in remote and online learning environments. One significant approach to developing ITSs is to leverage various statistical features to perform learning analytics activities and performance prediction [13], [14], and many previous studies have focused on engagement and dropout prevention, such as leveraging students' facial expressions for emotion recognition and engagement prediction [15]. On the other hand, as an advanced form of ITSs, conversational ITSs have been extensively investigated as educational dialogue systems [16]–[18], as they can provide adaptive instructions and real-time feedback to students. Most existing studies focus on learning the pedagogical strategies to teach the students of the given exercises [19], [20], or generating high-quality responses in the tutoring dialogues [21]. Latest studies [7], [22], [23] on interactive ITSs powered by LLMs have showcased the exceptional capabilities of natural language interactions.

B. Scaffolding in Children's Language Learning

Since the late 1970s, scaffolding has gained increasing popularity across various educational fields, especially in language learning contexts. This popularity is due to the crucial roles of the meaning-making process and linguistic assistance in students' language development [24], [25]. Teachers can apply scaffolding strategies, such as questioning, reformulation, repetition, and elaboration to assist English language learners in co-constructing content knowledge, thereby making these processes "visible" to them [10]. With the support and guidance of teachers, students are more likely to complete the given task and face similar challenges in the future with greater confidence [26], [27]. Previous research has identified several key characteristics essential for effective scaffolding [28]. The most salient feature is contingency. Teachers assess students' competency levels and dynamically adapt scaffolding strategies based on the learners' understanding and actions [24]. Another aspect of scaffolding is fading [29], [30]. In this process, teachers gradually withdraw the scaffolding as students are able to carry out tasks independently [31]. Thus, scaffolding is a temporary and adjustable process, with support aligned towards facilitating students' learning goals.

While research on scaffolding has enriched the understanding of teaching practices, the process is often limited to either one-on-one or one-to-many teacher-led instruction. This can result in limited access and fewer opportunities for students to engage in practice. Consequently, students might have fewer chances of being heard, scaffolded, and receiving feedback. Such limitations could adversely affect their language learning and usage. The advent of LLMs has empowered ITSs to provide scaffolding strategies in supporting students learning outside of classrooms in various disciplines, and they can make online learning personalized and more accessible.

III. Multi-modal Tutoring Systems with Pedagogical Instructions

A. Tutoring Language Learning via Image Description

Teaching and improving primary students' language learning through image description is a dynamic and engaging approach [32]. As the example shown in Figure 1, a learning session usually begins by presenting an image and encouraging students to observe it closely, then the teacher asks open-ended questions to stimulate their thinking, such as "*What do you see happening in this picture?*" or "*Can you describe the people or animals you see?*"

Beyond merely listing the objects in the image, the teacher further guides students to describe how things look, feel, or sound, and encourages students to use adjectives and adverbs. This exercise enhances their language skills including vocabulary, organization, and fluency [33]. To further develop their

TABLE I
TABLE OF THE DESCRIPTION OF PEDAGOGICAL INSTRUCTIONS.

Theory Type	Definition	Pedagogical Strategy
Knowledge Construction (Sullivan Palincsar, 1998)	The effortful, situated, and reflective process by which students solve problems and construct an understanding of concepts, phenomena, and situations.	Consistently assisting students in building upon their prior knowledge, organizing and synthesizing information, integrating ideas, and making inferences.
Inquiry-based Learning (Pedaste et al., 2015)	Engaging learners by creating real-world connections through questioning and exploration.	Guiding learners with explicit learning goals and helping them develop an explanatory learning process, breaking down complex tasks into small and manageable segments, making observations, asking questions, posing hypotheses, investigating, interpreting, and discussing.
Dialogic Teaching (Alexander, 2006)	The ongoing process of dialogue in stimulating and developing students' thinking, learning and understanding.	Co-constructing knowledge through dialogue and collaboration, encouraging the free exchange of ideas, using follow-up questions, clues, elaborations, reformulations, confirmations, or recaps, building on prior knowledge and understanding.
Zone of Proximal Development (Vygotsky, 1978)	The space between what a learner can do without assistance and what a learner can do with adult guidance or in collaboration with more capable peers.	Assessing the learner's current ability level, connecting content to learners' existing knowledge, breaking down a task into smaller, manageable components, and using prompts and cues to help students achieve a potential level beyond their current capabilities.

language skills, the teacher introduces new vocabulary related to the image, and engages students in storytelling. Students are asked to create a short story or summary, thereby stimulating their creativity and imagination to enrich the narrative with additional details and depth.

B. Multi-modal Systems as an Image Description Tutor

Beyond text-based interactive learning (e.g., math and coding tutoring), multi-modal capabilities are essential for building image description tutoring systems. Basically, it includes four functional aspects: vision modeling, speech recognition, natural language generation, and dialogue management. More specifically, given an image input, vision modeling is to capture the visual features and recognize various scenes, objects, and activities, as well as grounded knowledge such as spatial relationships [34]. Speech recognition is to convert student responses from audio to text. Natural language generation and dialogue management enable the tutoring system to interact with students via effective communication, including generating fluent, coherent, and descriptive language of the image, raising contextualized questions, and providing hints and explanations. In oral courses, spoken language assessment is also an integral component within machine-aided language learning, which is used to evaluate oral proficiency [35].

Upon the versatility and capability of LLMs, one can build tutoring systems without massive supervision from time-consuming manual annotation [19], [20]. Therefore, we leverage GPT-4V as an image description tutoring agent for language learning, since it is a multi-modal model that supports all four functional aspects, and shows state-of-the-art instruction-following and reasoning capabilities [36].

C. Enhancing Tutoring Systems with Pedagogical Instructions

In practical settings, teachers adhere to established pedagogical principles to enhance their instructional methods, demonstrating the efficacy of a more focused and systematic approach [37]. Consequently, to develop an image-based tutoring system that effectively motivates and supports students in language acquisition, we explore the impact of incorporating structured pedagogical strategies. On the other hand, LLMs are capable of following complex and detailed prompts, and performing as task-specialized agents [38]. Previous work shows that prompting in a structured manner is beneficial for complex instruction following [39], thus we split it into three parts: role & task definition, pedagogical instruction, and behavior constraint. Here is one template:

[Role & Task Definition] *You are a primary school language teacher who teaches me to describe the picture.*
[Pedagogical Instruction] *You are using the knowledge construction approach. This involves any one of the following: building on prior knowledge, selecting information, organizing information, integrating ideas, making inferences, and helping me describe the picture.*
[Behavior Constraint] *Ask me only one question at a time. Always wait for my input before proceeding to the next step. Correct my answers if they are inaccurate.*

Moreover, for the pedagogical instruction, we apply four constructivist learning theories (as shown in Table I) and conduct experiments on how they affect the scaffolding of language learning via image description.

Knowledge Construction When individuals encounter new information, they rely on their prior knowledge and personal experience to interpret it [40]. During this meaning-making process, learners reformulate the new information or restructure their existing knowledge, thereby achieving deeper understanding [41]. Prior research found that knowledge construction can range from simple restatements and paraphrasing to more complex activities like explanations, inferences, justifications, hypotheses, and speculations [42]. Specifically, teachers facilitate students' knowledge construction process by consistently assisting students in building upon their prior knowledge, organizing and synthesizing information, integrating ideas, and making inferences [43].

Inquiry-based Learning is a pedagogical approach that engages learners by creating real-world connections through questioning and exploration [44]. It aims to inspire students to take ownership of their learning journey. To support these in-

TABLE II
TABLE OF THE RUBRIC DEFINITION FOR EVALUATING SCAFFOLDING EFFICACY.

Dimension		Definition	Utterance Example
Cognitive Scaffolding	Feeding back	The teacher directly evaluates the behavior or response of the student.	Yes, the girl does look happy! Great! You're right.
	Hints	The teacher gives an explicit hint with respect to the expected answer.	Does he look happy, surprised, or something else? Look at the TV in the picture for a clue.
	Instructing	The teacher provides information so that the student knows what to do or how to do it. Request for a specific action (e.g., look at sth. or focus sth.).	Look at the things around them for clues. Remember to include what you've noticed about cleaning and organizing.
	Explaining	The teacher provides detailed information on "why" or clarification.	When someone opens their mouth like that and has tears on their face, it often does indicate that they are crying or upset.
	Modeling	The teacher demonstrates behavior (verbal or non-verbal) for imitation.	Just a small grammar tip: when we say "with the girl is dancing," we don't need the word "is" after "girl".
	Questioning	The teacher asks the student questions that require an active linguistic and cognitive answer.	Can you tell me if it's daytime or nighttime? And how can you tell?
Social-emotional Support		Responses related to emotion and motivation such as positive affirmation, showing empathy, promoting self-efficacy, fostering a sense of connectedness, encouraging perseverance, and other related constructs.	No problem at all! No worries, let's observe together!

quiry outcomes, researchers have proposed several scaffolding strategies [45], [46]. For example, teachers guide learners with explicit learning goals and help them develop an explanatory learning process [47]. Specifically, tasks are structured to minimize cognitive overload. Teachers break down complex tasks into small and manageable segments. This approach narrows the "problem space", enabling learners to focus their efforts and utilize available resources or tools effectively [46].

Dialogic Teaching highlights the role of talk in stimulating and developing students' thinking, learning, and understanding [48], [49]. To facilitate productive interactions, teachers encourage students by posing thought-provoking questions and inviting them to share their knowledge and experience, which aims not just to seek right answers, but also to elicit reasons and explanations [50]. This often involves a scaffolded dialogue pattern known as initiation-response-feedback pattern (IRF) [37]. Specifically, the teacher initiates a topic, students respond, and then obtain feedback [51]. In this cyclic IRF sequence, teachers guide students by using follow-up questions, clues, elaborations, reformulations, confirmations, or recaps, thereby maintaining students' active participation [52].

Zone of Proximal Development (ZPD) is defined as the space between what a learner can do without assistance and what a learner can do with adult guidance or in collaboration with more capable peers [9]. In Vygotsky's view, a learner's ability to bridge this gap between actual performance and potential ability depends on the scaffolding provided by more capable others [53]. In pedagogical contexts, scaffolding techniques involve several processes: assessing the learner's current ability level, connecting content to learners' existing knowledge, breaking down a task into smaller, manageable components, and using prompts and cues to help students achieve a potential level beyond their current capabilities.

Based on the aforementioned pedagogical theories, we summarized key features of each theory and synthesized corresponding instructional strategies for scaffolding within our ITS, as shown in Table I. These strategies, embedded in student-ITS conversational interactions, aim to provide guidance and support in image description tasks, thereby enhancing the overall learning experience.

IV. EVALUATING TUTORING SYSTEMS FROM SCAFFOLDING PERSPECTIVE

A. Building Student Capability Levels for Evaluation

Scaffolding strategies effectively support the learning processes of students. However, the needs, learning styles, and educational experiences of low- and high-achieving learners differ significantly [54]. First, low achievers often feel uncomfortable expressing their ideas because they may lack prior knowledge and self-confidence. They tend to wait for assistance rather than attempting to solve problems independently [55], [56]. Second, students with lower performance frequently encounter more misconceptions resulting in the need for more individualized learning paths and more interactive and adaptive scaffolds [57], [58].

These differences have led to the classification of students into *high-* and *low-ability* groups. In this study, *high-ability* students are defined as those with high language proficiency who can answer each question correctly with minimal support and guidance. Conversely, *low-ability* students are characterized by low language proficiency and are not able to answer questions independently. This classification serves two main purposes: first, to investigate how scaffolding strategies are applied to students with varying abilities, and second, to explore the differences among different tutoring systems.

B. Building Scaffolding Evaluation Rubric

When comparing the learning efficacy of tutoring systems, it is common to conduct assessments on post-learning performance, dropout rate, and user feedback. However, the scaffolding strategy in conversational interactions, another important aspect, is overlooked [59]. To evaluate the effectiveness of scaffolding strategies, here we introduce and refine a rubric of

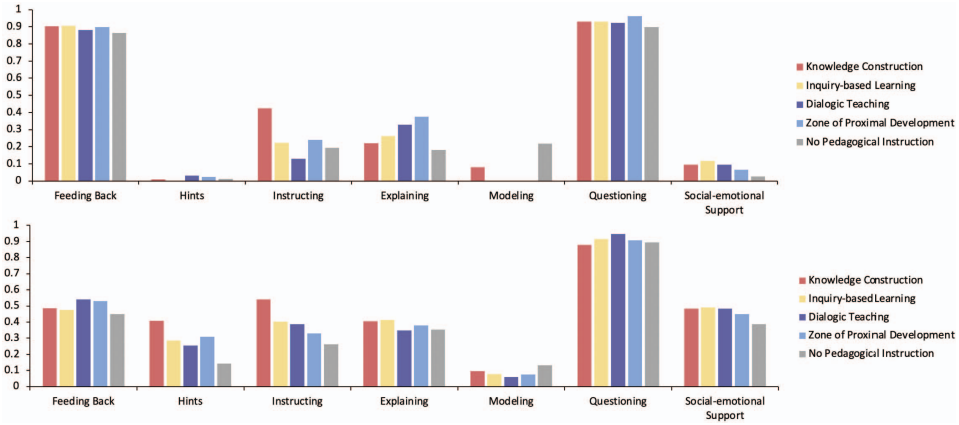

Fig. 2. Coding result of the five systems with different pedagogical instructions (Up: high-ability group; Down: low-ability group).

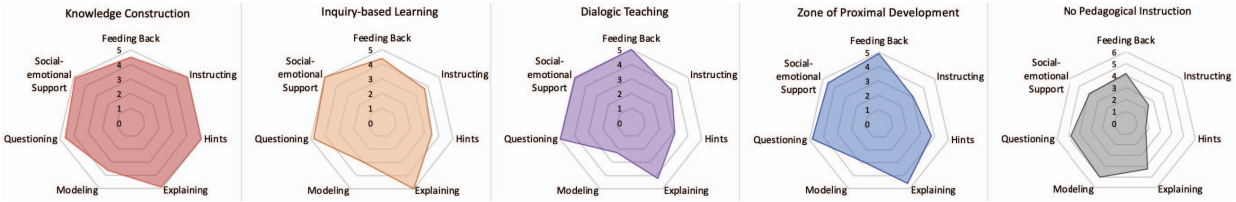

Fig. 3. Normalized capability scoring in seven dimensions of the five systems with different pedagogical instructions.

dialogic analysis at the utterance level. Based on previous pedagogical work [12], [37], our rubric is designed to understand how scaffolding is performed during students' language learning, and it consists of seven dimensions (as shown in Table II): Feeding back, Hints, Instructing, Explaining, Modeling, Questioning, and Social-emotional Support.

C. Experimental Setting and Qualitative Analysis

Inspired by real-world learning materials for primary school level 1 and level 2 students, we constructed a seed dataset for qualitative analysis. To improve the diversity of visual and language features, the selected 10 image description samples cover various scenes (e.g., classroom, playground, home), objects (e.g., family, kids, teacher), and activities (e.g., sports, reading). We simulate the learning process via human-machine interaction, where the tutoring system leads the conversation, and we feed user responses according to the assigned student group. The average turn number of each conversation is 22.5, we repeat each session across 5 pedagogical instruction types as well as 2 student groups (a system without pedagogical instruction is added as control), and the total collected utterance number is 2250. In our preliminary analysis, we observe that the system without any pedagogical instruction is also capable of utilizing visual features and organizing the conversation. For instance, when students get confused, the tutor will ask them to look at one specific part of the picture and encourage their attention to certain objects.

Based on our rubric, we code the system utterances and calculate the scores among five types of pedagogical instructions and between two students' ability levels. For each system utterance, a score of 1 was assigned if it corresponded with the dimensions of scaffolding strategies. Two linguistic annotators coded system generations independently, and we conducted two rounds of preliminary coding to consolidate the rubric description and reduce discrepancies. The final inter-annotator Cohen's Kappa is 0.75.

D. Experimental Results and Analysis

1) Comparison between high- and low-ability students: The concept of contingency emphasized the malleable feature of scaffolding in relation to students' understanding. Contingent support suggests that the tutor amplifies the level of support in reaction to student failure or diminishes it following student success. In this study, we compared the scaffolding strategies that ITS applied to *high-* and *low- ability* students. We observed that systems with four pedagogical instructions outperformed in providing contingent support, managing to increase the degree of contingency for *low- ability* students while reducing it for those with high abilities. This increased level of contingency was associated with a rise in the provision of hints, instruction, explanations, modeling, and social-emotional support.

Specifically, in Figure 2, we observed that when engaging with *high-ability* students, the scaffolding typically begins with positive affirmation, followed by guiding the students through questions and clarifying their answers. In contrast, when interacting with *low-ability* students, systems tend to pay more attention to social-emotional support. Additionally, the scaffolding for these students includes more hints and expla-

TABLE III
EXPERIMENTAL RESULTS OF AUTOMATED SCORING BY LEVERAGING LLMS AS EVALUATOR.

	Zero-Shot Inference		1-Shot Inference		3-Shot Inference	
Model	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**	**Accuracy**	**F1 Score**
LLaMA-2-7B-Chat	0.533	0.497	0.548	0.536	0.654	0.644
LLaMA-2-13B-Chat	0.599	0.375	0.609	0.583	0.708	0.698
Vicuna-13B-V1.5	0.417	0.368	0.670	0.650	0.765	0.757
Mistral-7B-Instruct-V0.1	0.480	0.426	0.743	0.733	0.777	0.769
Zephyr-7B-Beta	0.711	0.706	0.746	0.732	0.785	0.778
GPT-3.5-turbo-1106	0.783	0.764	0.787	0.775	0.805	0.795

nations, and in most cases, they provide structured examples for the students to imitate.

2) Comparison among Scaffolding Tutors: Figure 3 illustrates capability scores in 7 dimensions of various scaffolding strategies (normalized by the max value of each dimension). Those equipped with pedagogical instructions outperformed the ones without pedagogical instruction in each dimension except for *Modeling*. This can be attributed to the difference of *Modeling* contents. The system lacking pedagogical instructions predominantly offers direct answers to students who are unable to answer questions themselves. In contrast, other ones, particularly when interacting with students of lower language proficiency, initiate with hints and explanations, aiming to encourage and assist the students.

Hints are one of the most supportive dimensions in facilitating students with language learning [60]. During the scaffolding process, hints work as moderators between students and knowledge. They effectively help learners access contextual information and important shortcuts, ultimately assisting students in language development [61]. In this study, the occurrence of hints is significantly more prevalent among pedagogical-based tutoring systems, particularly in interactions with *low-ability* students due to their limited vocabulary and lower language proficiency. This suggests that they are capable of dynamically adapting their scaffolding approach to meet the diverse needs of learners, prioritizing language support where it is most needed.

Language learning occurs through imitation, reinforcement of contextual or verbal stimuli, practice of correct responses, and immediate correction of incorrect responses by the teacher [62]. Additionally, several strategies have been identified for introducing or explaining new topics, concepts, or terms to students, including explaining, reformulating, clarifying, and exemplifying [63]. These approaches effectively build connections between new information and students' prior knowledge. In this study, we observed that tutors with pedagogical instructions are capable of explaining, describing, elaborating, and comparing new knowledge by leveraging students' familiar concepts. They could also provide grammatical structures, contextual forms, or examples for correction, thus enabling students to construct descriptions accurately. Conversely, the tutor without pedagogical instructions often provides direct answers with less personalized scaffolding and fewer interactive learning opportunities, leading to a more passive and less engaging learning experience for students.

V. TOWARDS AUTOMATED SCAFFOLDING EVALUATION

A. Leveraging LLMs for Automated Evaluation

Since the student scaffolding can be significantly affected by the pedagogical tutor, in practical use cases, transforming from manual to automated utterance scoring represents a significant advancement in scalable, accurate, and efficient evaluation. In this section, we investigate the potential of employing LLMs for automated scoring. Recent work shows that LLMs can achieve a high correlation with human judgments on various tasks [64]. Based on previous work, we designed a natural language instruction for the scoring task according to our rubric. The prompt is created by concatenating the scoring criteria, and utterance context, and then fed to a model for prediction. The output is the labeled types for each dimension based on the defined schema.

B. Experimental Results and Analysis

To demonstrate the feasibility and efficacy of automated evaluation and compare the performance of LLMs, we use our manual annotation as a reference, and results are presented in terms of correlation with human judgments, using accuracy and F1 scores. Here we selected and tested a list of representative models. As shown in Table III, while most models cannot provide reasonable results under zero-shot inference setting, the performance can be significantly improved by adding only 3 examples (i.e., 3-shot inference). In particular, for LLaMA-based models (i.e., LLaMA-2 and Vicuna), a larger parameter size (13B vs. 7B) brings higher accuracy. The 7B models can achieve state-of-the-art performance (e.g., Mistral-7B, Zephyr-7B), and are comparable to GPT-3.5 in the few-shot setting. This demonstrates the feasibility of utilizing open LLMs to build automated and scalable scaffolding benchmarks.

VI. CONCLUSION

Our work contributes to an in-depth understanding of LLM-based ITSs from the scaffolding perspective. First, we built a tutoring system that guides children to describe images for language learning, and enhanced it with pedagogical instructions. Second, we developed and validated a seven-dimension rubric to assess scaffolding strategies for different student groups. Our findings offer valuable insights into instructional design, improving the learning experience through interactive and supportive scaffolding strategies, which are aligned with personalized learning needs. Third, we leveraged LLMs to automate the scaffolding evaluation framework, paving the way to benchmark various conversational tutoring systems.

REFERENCES

[1] E. Mousavinasab, N. Zarifsanaiey, S. R. Niakan Kalhori, M. Rakhshan, L. Keikha, and M. Ghazi Saeedi, "Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods," Interact. Learn. Environ., vol. 29, no. 1, pp. 142–163, Jan. 2021, doi: 10.1080/10494820.2018.1558257.

[2] A. Keleş, R. Ocak, A. Keleş, and A. Gülcü, "ZOSMAT: Web-based intelligent tutoring system for teaching–learning process," Expert Syst. Appl., vol. 36, no. 2, pp. 1229–1239, Mar. 2009, doi: 10.1016/j.eswa.2007.11.064.

[3] J. A. Kulik and J. D. Fletcher, "Effectiveness of Intelligent Tutoring Systems," Rev. Educ. Res., vol. 86, no. 1, pp. 42–78, Mar. 2016, doi: 10.3102/0034654315581420.

[4] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, and E. H. Chi, "Emergent Abilities of Large Language Models," Transactions on Machine Learning Research, 2022.

[5] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, and J. Schulman, "Training language models to follow instructions with human feedback," Adv. Neural Inf. Process. Syst., vol. 35, pp. 27730–27744, 2022.

[6] Y. Chen, N. Ding, H.-T. Zheng, Z. Liu, M. Sun, and B. Zhou, "Empowering Private Tutoring by Chaining Large Language Models," 2023, [Online]. Available: http://arxiv.org/abs/2309.08112.

[7] B. D. Nye, D. Mee, and M. G. Core, "Generative Large Language Models for Dialog-Based Tutoring: An Early Consideration of Opportunities and Concerns," CEUR Workshop Proc., vol. 3487, pp. 78–88, 2023.

[8] E. Kasneci, K. Seßler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, and S. Krusche, "ChatGPT for good? On opportunities and challenges of large language models for education," Learn. Individ. Differ., vol. 103, p. 102274, 2023.

[9] L. S. Vygotsky, Mind in Society. Harvard University Press, 1978.

[10] P. Gibbons, Scaffolding language, scaffolding learning. Heinemann Portsmouth, NH, 2002.

[11] J. Hammond, Scaffolding: Teaching and learning in language and literacy education. ERIC, 2001.

[12] J. van de Pol, M. Volman, and J. Beishuizen, Scaffolding in Teacher–Student Interaction: A Decade of Research, Educ. Psychol. Rev., vol. 22, no. 3, pp. 271–296, Sep. 2010, doi: 10.1007/s10648-010-9127-6.

[13] J. X. Weng, A. Y. Huang, O. H. Lu, I. Y. Chen, and S. J. Yang, "The implementation of precision education for learning analytics," in 2020 IEEE Int. Conf. on Teaching, Assessment, and Learning for Engineering (TALE), Dec. 2020, pp. 327–332.

[14] F. Ouyang, M. Wu, L. Zheng, L. Zhang, and P. Jiao, "Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course," Int. J. Educ. Technol. High. Educ., vol. 20, no. 1, pp. 1–23, 2023.

[15] D. Leony, P. J. Muñoz-Merino, A. Pardo, and C. D. Kloos, "Provision of awareness of learners' emotions through visualizations in a computer interaction-based environment," Expert Syst. Appl., vol. 40, no. 13, pp. 5093–5100, 2013.

[16] J. Macina, N. Daheim, L. Wang, T. Sinha, M. Kapur, I. Gurevych, and M. Sachan, "Opportunities and Challenges in Neural Dialog Tutoring," in Proc. of the 17th Conf. of the Eur. Chapter of the Assoc. for Comput. Linguistics, May 2023, pp. 2349–2364.

[17] S. Ruan, L. Jiang, J. Xu, B. J. K. Tham, Z. Qiu, Y. Zhu, E. L. Murnane, E. Brunskill, and J. A. Landay, "Quizbot: A dialogue-based adaptive learning system for factual knowledge," in Proc. of the 2019 CHI Conf. on Human Factors in Computing Systems, May 2019, pp. 1–13.

[18] S. Wollny, J. Schneider, D. Di Mitri, J. Weidlich, M. Rittberger, and H. Drachsler, "Are we there yet?-a systematic literature review on chatbots in education," Front. Artif. Intell., vol. 4, p. 654924, 2021.

[19] K. Stasaski, K. Kao, and M. A. Hearst, "CIMA: A large open access dialogue dataset for tutoring," in Proc. of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications, July 2020, pp. 52–64.

[20] A. Suresh, J. Jacobs, C. Harty, M. Perkoff, J. H. Martin, and T. Sumner, "The talkmoves dataset: K-12 mathematics lesson transcripts annotated for teacher and student discursive moves," arXiv preprint arXiv:2204.09652, 2022.

[21] L. Wang, M. Sachan, X. Zeng, and K. F. Wong, "Strategize Before Teaching: A Conversational Tutoring System with Pedagogy Self-Distillation," arXiv preprint arXiv:2302.13496, 2023.

[22] Y. Chen, N. Ding, H. T. Zheng, Z. Liu, M. Sun, and B. Zhou, "Empowering Private Tutoring by Chaining Large Language Models," arXiv preprint arXiv:2309.08112, 2023.

[23] Y. Dan, Z. Lei, Y. Gu, Y. Li, J. Yin, J. Lin, L. Ye, Z. Tie, Y. Zhou, Y. Wang, and A. Zhou, "Educhat: A large-scale language model-based chatbot system for intelligent education," arXiv preprint arXiv:2308.02773, 2023.

[24] A. Walqui, "Scaffolding instruction for english language learners: A conceptual framework," Int. J. Biling. Educ. Biling., vol. 9, no. 2, pp. 159–180, 2006, doi: 10.1080/13670050608668639.

[25] H. Kayi-Aydar, "Scaffolding language learning in an academic ESL classroom," ELT J., vol. 67, no. 3, pp. 324–335, Jul. 2013, doi: 10.1093/elt/cct016.

[26] L. C. de Oliveira, L. Jones, and S. L. Smith, "Interactional scaffolding in a first-grade classroom through the teaching–learning cycle," Int. J. Biling. Educ. Biling., vol. 26, no. 3, pp. 270–288, Mar. 2023, doi: 10.1080/13670050.2020.1798867.

[27] M. Y. Damanhouri, "The Effectiveness of Scaffolding as a Teaching Strategy in Enhancing English Language Learners' Motivation in Writing: A Case Study," J. Arts Humanit., vol. 10, no. 03, pp. 49–58, 2021, doi: 10.18533/jah.v10i03.2056.

[28] T. Gonulal and S. Loewen, "Scaffolding Technique," in The TESOL Encyclopedia of English Language Teaching, Wiley, 2018, pp. 1–5.

[29] S. Puntambekar, "Distributed Scaffolding: Scaffolding Students in Classroom Environments," Educ. Psychol. Rev., vol. 34, no. 1, pp. 451–472, 2022, doi: 10.1007/s10648-021-09636-3.

[30] S. Puntambekar and R. Hubscher, "Tools for Scaffolding Students in a Complex Learning Environment: What Have We Gained and What Have We Missed?," Educ. Psychol., vol. 40, no. 1, pp. 1–12, Mar. 2005, doi: 10.1207/s15326985ep4001-1.

[31] S. P. Lajoie, "Extending the Scaffolding Metaphor," Instr. Sci., vol. 33, no. 5, pp. 541–557, 2005, doi: 10.1007/s11251-005-1279-2.

[32] A. Wright, Pictures for language learning. Cambridge University Press, 1989.

[33] A. Pinter, Teaching young language learners. Oxford University Press, 2017.

[34] Y. Yamada, Y. Bao, A. K. Lampinen, J. Kasai, and I. Yildirim, "Evaluating Spatial Understanding of Large Language Models," arXiv preprint arXiv:2310.14540, 2023.

[35] J. H. Wong, H. Zhang, and N. F. Chen, "Variations of multi-task learning for spoken language assessment," in Proc. Interspeech, 2022, pp. 4456–4460.

[36] Z. Yang, L. Li, K. Lin, J. Wang, C. C. Lin, Z. Liu, and L. Wang, "The dawn of lmms: Preliminary explorations with gpt-4v (ision)," arXiv preprint arXiv:2309.17421, vol. 9, no. 1, 2023.

[37] G. Wells, Dialogic inquiry: Towards a sociocultural practice and theory of education. New York, NY, US: Cambridge University Press, 1999.

[38] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, and J. Schulman, "Training language models to follow instructions with human feedback," Adv. Neural Inf. Process. Syst., vol. 35, pp. 27730–27744, Dec. 2022.

[39] T. Khot, H. Trivedi, M. Finlayson, Y. Fu, K. Richardson, P. Clark, and A. Sabharwal, "Decomposed Prompting: A Modular Approach for Solving Complex Tasks," arXiv preprint arXiv:2210.02406, 2022.

[40] L. B. Resnick and S. N. R. C. (US). C. on R. in Mathematics, Education and learning to think. National Academy Press Washington, DC, 1987.

[41] A. S. Palincsar, "Social constructivist perspectives on teaching and learning," Annu. Rev. Psychol., vol. 49, pp. 345–375, 1998, doi: 10.1146/annurev.psych.49.1.345.

[42] C. K. K. Chan, P. J. Burtis, M. Scardamalia, and C. Bereiter, "Constructive Activity in Learning From Text," Am. Educ. Res. J., vol. 29, no. 1, pp. 97–118, Mar. 1992, doi: 10.3102/00028312029001097.

[43] J. van Aalst, "Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses," Int. J. Comput. Collab. Learn., vol. 4, no. 3, pp. 259–287, Sep. 2009, doi: 10.1007/s11412-009-9069-5.

[44] M. Pedaste et al., "Phases of inquiry-based learning: Definitions and the inquiry cycle," Educ. Res. Rev., vol. 14, pp. 47–61, Feb. 2015, doi: 10.1016/j.edurev.2015.02.003.

[45] K. L. McNeill, D. J. Lizotte, J. Krajcik, and R. W. Marx, "Supporting Students' Construction of Scientific Explanations by Fading Scaffolds

in Instructional Materials," J. Learn. Sci., vol. 15, no. 2, pp. 153–191, Dec. 2006.
[46] B. J. Reiser, "Scaffolding Complex Learning: The Mechanisms of Structuring and Problematizing Student Work," J. Learn. Sci., vol. 13, no. 3, pp. 273–304, Jul. 2004, doi: 10.1207/s15327809jls1303-2.
[47] Y. S. Hsu, T. L. Lai, and W. H. Hsu, "A Design Model of Distributed Scaffolding for Inquiry-Based Learning," Res. Sci. Educ., vol. 45, no. 2, pp. 241–273, 2015, doi: 10.1007/s11165-014-9421-2.
[48] R. Alexander, Culture and pedagogy. Oxford, UK, 2000.
[49] R. Alexander, Education as Dialogue: Moral and Pedagogical Choices for a Runaway World. Hong Kong Institute of Education, 2006.
[50] N. Mercer and C. Howe, "Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory," Learn. Cult. Soc. Interact., vol. 1, no. 1, pp. 12–21, Mar. 2012, doi: 10.1016/j.lcsi.2012.03.001.
[51] G. Wells and R. M. Arauz, "Dialogue in the Classroom," J. Learn. Sci., vol. 15, no. 3, pp. 379–428, Jul. 2006, doi: 10.1207/s15327809jls1503-3.
[52] R. Alexander, "Developing dialogic teaching: Genesis, process, trial," Res. Pap. Educ., vol. 33, no. 5, pp. 561–598, 2018.
[53] E. B. Raymond, Learners with mild disabilities : a characteristics approach. Allyn and Bacon Boston, 2000.
[54] C. H. Hargis, Teaching low achieving and disadvantaged students. Charles C Thomas Publisher, 2006.
[55] S. Haruehansawasin and P. Kiattikomol, "Scaffolding in problem-based learning for low-achieving learners," J. Educ. Res., vol. 111, no. 3, pp. 363–370, May 2018, doi: 10.1080/00220671.2017.1287045.
[56] A. Soller, B. Goodman, F. Linton, and R. Gaimari, "Promoting Effective Peer Interaction in an Intelligent Collaborative Learning System," in Intelligent Tutoring Systems, B. P. Goettl, H. M. Halff, C. L. Redfield, and V. J. Shute, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, pp. 186–195.
[57] J. Sweller, "Cognitive Load Theory," in The psychology of learning and motivation: Cognition in education, Vol. 55, San Diego, CA, US: Elsevier Academic Press, 2011, pp. 37–76.
[58] F. Reinhold, S. Hoch, B. Werner, J. Richter-Gebert, and K. Reiss, "Learning fractions with and without educational technology: What matters for high-achieving and low-achieving students?," Learn. Instr., vol. 65, p. 101264, Feb. 2020, doi: 10.1016/j.learninstruc.2019.101264.
[59] M. C. Duffy and R. Azevedo, "Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system," Comput. Hum. Behav., vol. 52, pp. 338–348, 2015.
[60] M. Celce-Murcia and E. Olshtain, Discourse and context in language teaching: A guide for language teachers. Cambridge University Press, 2000.
[61] S. Khaliliaqdam, "ZPD, Scaffolding and Basic Speech Development in EFL Context," Procedia - Soc. Behav. Sci., vol. 98, pp. 891–897, May 2014, doi: 10.1016/j.sbspro.2014.03.497.
[62] H. C. Dulay and M. K. Burt, "SHOULD WE TEACH CHILDREN SYNTAX?," Lang. Learn., vol. 23, no. 2, pp. 245–258, Dec. 1973, doi: 10.1111/j.1467-1770.1973.tb00659.x.
[63] J. C. Richards and T. S. Rodgers, Approaches and Methods in Language Teaching, 3rd ed. Cambridge: Cambridge University Press, 2014.
[64] M. Li, T. Shi, C. Ziems, M. Y. Kan, N. Chen, Z. Liu, and D. Yang, "CoAnnotating: Uncertainty-Guided Work Allocation between Human and Large Language Models for Data Annotation," in Proc. of the 2023 Conf. on Empirical Methods in Natural Language Processing, Dec. 2023, pp. 1487–1505.