Beyond the Grey Area: Exploring the Effectiveness of Scaffolding as a Learning Measure

Nadine Schulze Bernd[1] and Irene-Angelica Chounta[1][0000-0001-9159-0664]

Department of Human-Centered Computing and Cognitive Science, University of Duisburg-Essen, Duisburg, Germany
nadine.schulze-bernd@stud.uni-due.de, irene-angelica.chounta@uni-due.de

Abstract. In this paper, we aim to explore students' help-seeking patterns, and learning performance using a computational model of the Zone of Proximal Development (ZPD). We used student traces in four courses supported by Intelligent Tutoring Systems (ITSs) to assess whether a student may be in their ZPD. Then, we analyzed students' help-seeking behavior and their performance after receiving scaffolding to investigate if and under which circumstances our assessment may act as a proxy of the ZPD. Results suggest that the computational model may offer an acceptable approximation of the student's ZPD. However, several factors, such as the difficulty of the learning task, the student's age, and their attitude regarding hints, may affect modeling students' cognitive states. We discuss the implications of this research on the design of computational student models for providing personalized and adaptive feedback and scaffolding. We envision that this work contributes toward bridging the gap between theory and practice for AI-assisted tutoring and scaffolding, and it provides insights regarding evaluation techniques of theory-driven, computational approaches.

Keywords: Intelligent Tutoring Systems · Student Modeling · Zone of Proximal Development · Scaffolding · Personalization · Adaptation.

1 Introduction

Intelligent Tutoring Systems (ITSs) aim to model students' cognitive and knowledge state to provide tailored content and instruction that address the learners' needs, to appropriately challenge and support them [9, 6]. Related work suggests that ITSs can support teaching and learning to that end, but there are still open challenges to overcome. For example, ITSs might force the student to engage in extensive discussions about material they have already mastered instead of addressing content they need help with [11, 13] causing boredom or frustration, which are both predictors of poor learning gains [8]. This risk can be addressed by using established pedagogical theories to drive the adaptation of instruction and content to meet the learners' needs. The Zone of Proximal Development (ZPD) [25] is a prominent pedagogical theory that refers to the

difference between what a learner can do independently and what they can do with assistance. Teaching within this zone is considered to be most effective and efficient as it enables challenging students without causing frustration or loss of motivation [4].
There have been attempts to model ZPD in ITSs and adjust pedagogical interventions accordingly [15, 20]. [5] attempted to model the ZPD based on the system's predictions about whether a student is able to answer a question correctly or not. This approach defines the **"Grey Area" (GA)** as the area in which the student model cannot predict with acceptable accuracy whether a student is able to answer the question correctly. If the model predicts a student to be in this area, this might indicate that the student's knowledge state falls within the ZPD and the student may need scaffolding to give a correct answer. If the model assesses that the student will be able to answer the question correctly, this might indicate that the student is above the ZPD and will be able to answer the question without help. If the model predicts that the student will answer incorrectly, then this might indicate that the student is below the ZPD and will not be able to answer the question even when scaffolding is available [5]. [11] provided preliminary results regarding this approach while testing it for the case of a dialogue-based tutor that supported K-12 students' conceptual knowledge about Physics. This paper aims to further validate the GA approach by a) expanding the GA's investigation over four different datasets and STEAM domains; and, b) exploring metrics that use students' help-seeking patterns and performance to assess the method's effectiveness as a proxy of the ZPD.
The ZPD is a theoretical concept that lacks an objective way to measure whether or when a student is in the ZPD or not. The number of correct and incorrect answers above, below, and within the GA can provide meaningful indications. However, these two metrics depend on the predictive accuracy of the student model and, therefore, cannot provide evidence regarding the effectiveness of the approach. At the same time, assistance is a significant concept of the ZPD [16, 20]. We argue that to validate whether a student is in their ZPD, one should examine the student's performance in relation to the help they received on specific tasks. In ITSs, assistance for solving a problem is mainly provided through hints. Thus, the usage of hints associated with the outcome in terms of performance – that is, whether a student carries out a learning step correctly or not after receiving a hint – can provide an indication for the student's ZPD [4, 20]. Here, we aim to answer the following research questions:
RQ1: Do students provide more correct answers when above the Grey Area than below?
RQ2: Do students provide more incorrect answers when below the Grey Area than above?
RQ3: Do students provide more correct answers after using a hint when they are within the GA than below?

In the following sections, we present research related to the concept of ZPD, ITSs, and existing approaches for modeling the ZPD. Then, we describe the methodological approach we followed to explore our questions, and we report the results. In section 5, we discuss our findings regarding theoretical and practical implications and we conclude with limitations of the study and plans for future work.

2 Related Work

2.1 The Zone of Proximal Development in Education

The Zone of Proximal Development (ZPD) was introduced by psychologist Lev Vygotsky and refers to the gap between learners' current level of ability and their potential level of development. It describes the difference between what a child can do independently and what they can do with the assistance of a "*more knowledgeable other*" [25]. When it comes to practice, there is no specification regarding the type of assistance that should be offered to learners or how it should be provided and withdrawn [16, 15]. In the literature, the concept of **scaffolding** is often associated with the ZPD [16]. While some researchers understand scaffolding as a direct application of teaching in the ZPD, others criticize it as a narrow way of operationalizing the ZPD as it only partially reflects the richness of this concept [23].

2.2 Approaches to implement the ZPD in ITSs

Operationalizations of the ZPD have been used in ITSs to optimize student problem-solving by adapting the level of difficulty and scaffolding appropriately. For example, [15] developed the ITS *Ecolab* with a ZPD operationalization to offer students help that was reflective of the complexity of the learning material. To that end, two additional concepts were introduced: the *Zone of Available Assistance* (ZAA) and the *Zone of Proximal Adjustment* (ZPA). The system assessed each learner's ZPA based on the levels the learner used at each step by using domain knowledge representations and Bayesian Belief Networks. Hence, the assistance to a learner was flexible and capable of being increased and decreased [15]. [20] extended this approach by introducing factors related to affect, such as boredom or frustration. The authors introduced the *Specific ZPD* (SZPD), an operationalized definition of the ZPD, which consists of a mastery criterion and a ZPD criterion. The mastery criterion addresses effectiveness by determining when the student can move on to the next unit. The ZPD criterion determines whether the student's learning was efficient which means that the student stays in a zone where they are neither too frustrated nor too bored.

Another approach to model the ZPD in ITSs is the *Grey Area* approach [5]. The GA describes the area in which a student model cannot predict with acceptable accuracy whether a student is able to answer a question correctly (see Figure 1). The authors proposed that if the model predicts a learner to be in this area, this might indicate that the learner's knowledge state falls within the ZPD. The space above the GA describes the region where the probability is considerably higher than the cutoff threshold (i.e., the student is predicted to answer correctly), and therefore, it may indicate the area above the ZPD in which the student is able to answer a question without assistance. The space below the GA describes the area with probabilities considerably lower than the cutoff threshold (i.e., the student is predicted to answer incorrectly) and, therefore, may indicate the area below the ZPD in which the student is not able to carry out the task even with assistance. For a first validation of the GA approach, [5] used data from students who interacted with a dialogue-based tutoring system called *Rimac*. A follow-up study [11] used the proposed approach to guide students who were practicing Rimac through adaptive lines of reasoning. The results indicated a positive impact of the student model on students' practicing times.

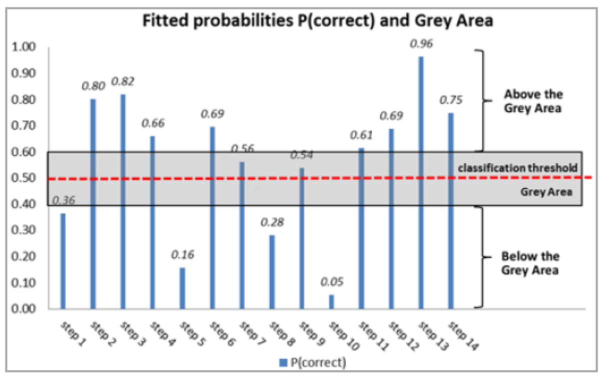

Fig. 1. Concept of the GA approach according to [5]

3 Methodology

3.1 Datasets

We used four datasets from students' interactions with different ITSs. The datasets were provided via Datashop [14]. DataShop is an open data repository that provides secure data storage of educational data, analysis and visualization tools. In particular, here we used the following datasets:

- **Fractions** dataset (Mathematics), ID 671. 77 students interacted with an ITS for 4th- and 5th-grade fractions learning [22]. In total, students performed 15269 steps practicing 20 KCs.
- **Genetics** dataset, ID 1195. The Genetics Cognitive Tutor supports genetics problem-solving for high school students who study biology [7]. 139 students interacted with the tutor and performed 128523 steps practicing 28 KCs.

– **Physics** dataset, ID 126. The dataset contains logs of student homework done in the ITS Andes for the General Physics course taught at the United States Naval Academy(USNA) [12]. In total 66 students performed 138958 steps while practicing 250 KCs.
– **Stoichiometry** dataset (Chemistry), ID 256. Stoichiometry is typically learned in the 10th or 11th grade in U.S. high schools and involves understanding basic chemistry concepts [17]. 418 students interacted with an ITS on worked examples and performed 124109 steps practicing 36 KCs.

Each dataset consists of students' transactions with an ITS to carry out a *problem*. A problem typically involves multiple *steps*. Each *step* is part of the solution to a problem. Before carrying out a step correctly, students may provide incorrect entries or ask for hints. In order to perform a step correctly, students need to know one or more specific skills or concepts, known as *knowledge components (KCs)*. The student's first attempt on a step is assessed using the Additive Factor Model (AFM) [3] to calculate a *predicted error rate*, ranging between 1 and 0; 1 suggests that a student's first attempt will be an error (incorrect attempt or hint request) and 0 that the first attempt will be correct.

3.2 Method Setup

We applied the GA approach based on the information provided in the four datasets, and as described by [5]. Then, we compared students' performance in and outside their Grey Areas to validate the initial assumption of the GA (RQ1, RQ2). Finally, we isolated the instances in students' practice, where a student received help from the ITS in the form of a hint. For the first hint on a step, we documented the student's follow-ups. This resulted in sequences of **hint-response** actions that characterized the practice of students. In particular, we identified three potential sequences: a) **hint-correct**: the student received a hint for a learning step and they proceeded to carry out this step correctly; b) **hint-incorrect**: the student received a hint for a learning step and they proceeded to carry out this step incorrectly; c) **hint-hint**: the student received a hint for a learning step and they proceeded to ask for another hint. To investigate RQ3, we calculated how many sequences of hint-correct, hint-incorrect, and hint-hint appear within and outside the GA.

The GA was operationalized as the area around the classification threshold of a binary classifier that would assess a student's attempt as correct or incorrect. For simplicity, the GA was modeled symmetrically around the cutoff threshold. The variable Predicted Error Rate was used to calculate the probability of correctness per step as $Prob(Correctness) = 1 - PredictedErrorRate$. The Predicted Error Rate was calculated for each step as a function of the student's proficiency, the difficulty of the KC, the learning rate and the student's prior practice for the KC. Steps with combined KCs were omitted from further analysis. To classify the students' attempts as correct or incorrect based on the Predicted Error Rate, we defined a probability cutoff (or else, optimal classification threshold, opt_t) to assign each step based on the predicted error rate. The classification threshold was calculated over each dataset (all students and all steps) using the receiver operating characteristic (ROC) curve with the aim of maximizing the sum of true positive and true negative rates. Finally, we defined the upper (1) and lower (2) boundaries of the GA as a function of opt_t to ensure that the GA would not extend beyond the range of the Predicted Error Rate.

$$GA_{upperboundary} = opt_t + ((1 - opt_t)/2) \tag{1}$$

$$GA_{lowerboundary} = opt_t - (opt_t/2) \tag{2}$$

To answer the research questions, we calculated for each dataset the following metrics: total number of steps per area (below, within, above), number of steps that were answered correctly (**RQ1**) or incorrectly (**RQ2**) for each of the three areas, and the number of hint-correct, hint-incorrect and hint-hint sequences per area (**RQ3**). Then, we conducted paired sample Wilcoxon Signed Rank Tests to investigate the statistical significance of our results. We applied the Bonferroni correction to account for multiple comparisons.

4 Results

4.1 Descriptive Analysis

Table 1 presents the descriptives regarding the students' performance for the first attempts. Three out of four datasets (Fractions, Physics and Stoichiometry) are characterized by more correct responses than incorrect ones. The only exception is the Genetics dataset, for which most responses were incorrect (54%). In all cases, less than 17% of hints were given out - for Fractions and Stoichiometry, only 4% of the responses corresponded to hints. The number of correct vs. incorrect responses in combination with the number of hints may be an indication of the difficulty of the learning task, suggesting that the content of the Genetics tutor was perceived as the most difficult among all four cases.

To answer RQ1 and RQ2, we calculated the distribution of correct and incorrect answers that students gave and the hints they received within the GA, and the areas above and below, for the four datasets under study (Table 2). Furthermore, we calculated the percentage ratio of correct, and incorrect answers and hints over the total number of answers per area (Figure 2). Our results suggest that the overall assumption regarding the correctness of answers within and outside the GA is confirmed. For all datasets, most correct answers fell above the GA (94% for Fractions, 78% for Genetics, 91% for Physics and 90% for the Stoichiometry datasets). Similarly, most incorrect answers were given in the area below GA (42% for Fractions, 56% for Genetics, 22% for Physics, and 53% for the Stoichiometry datasets). Hints were distributed over all three areas. However, the results suggest that hints are mainly requested below and within the GA, while fewer hints are requested above the GA. Next, we explored the help-seeking behavior of students over the GA and adjacent areas. To do so, we counted the number of sequences that involved hints over the different areas

Fractions	Correct	Hints	Incorrect
Number of steps	5634	309	2060
Percentage (%)	70	4	26
Genetics	**Correct**	**Hints**	**Incorrect**
Number of steps	1825	1104	3396
Percentage (%)	29	17	54
Physics	**Correct**	**Hints**	**Incorrect**
Number of steps	89851	15534	20011
Percentage (%)	72	12	16
Stoichiometry	**Correct**	**Hints**	**Incorrect**
	Correct Attempts	Hints	Incorrect Attempts
Number of steps	70188	3958	21749
Percentage (%)	73	4	23

Table 1. Descriptive statistics of first attempts' performance in terms of correctness.

	Below GA			Within GA			Above GA		
	c	h	i	c	h	i	c	h	i
Fractions	113	107	158	4051	200	1804	1470	2	98
Genetics	139	749	1135	1257	348	2145	429	7	116
Physics	1768	4613	1829	64134	10433	16245	23949	488	1937
Stoichiometry	517	225	827	4485	260	2039	1102	7	118

Table 2. The distribution of correct answers, hints, and incorrect answers within the Grey area and its adjacent areas, below and above.

under study. That is, we calculated the number of hint-correct, hint-hint, and hint-incorrect sequences that appeared in each of the three areas. The results are presented in Table 3. Figure 3 presents the percentage ratio of the help-seeking sequences over the total number of sequences in the three areas. The results showed that the ratio of hint-correct sequences is highest above the GA for all datasets. The ratio of hint-incorrect sequences appears to be distributed over all three areas while the ratio of hint-hint sequences is highest below the GA.

4.2 Inferential statistics

To answer **RQ1**, we calculated the number of correct answers above and below the GA for each student in the four datasets. Then, we compared the two groups using the Wilcoxon Signed Rank Test for paired samples. The results of this analysis are presented in Table 4. The results suggest that students provide more correct answers when above the GA than below. For three out of four datasets (Fractions, Physics, and Stoichiometry) these findings are statistically significant.

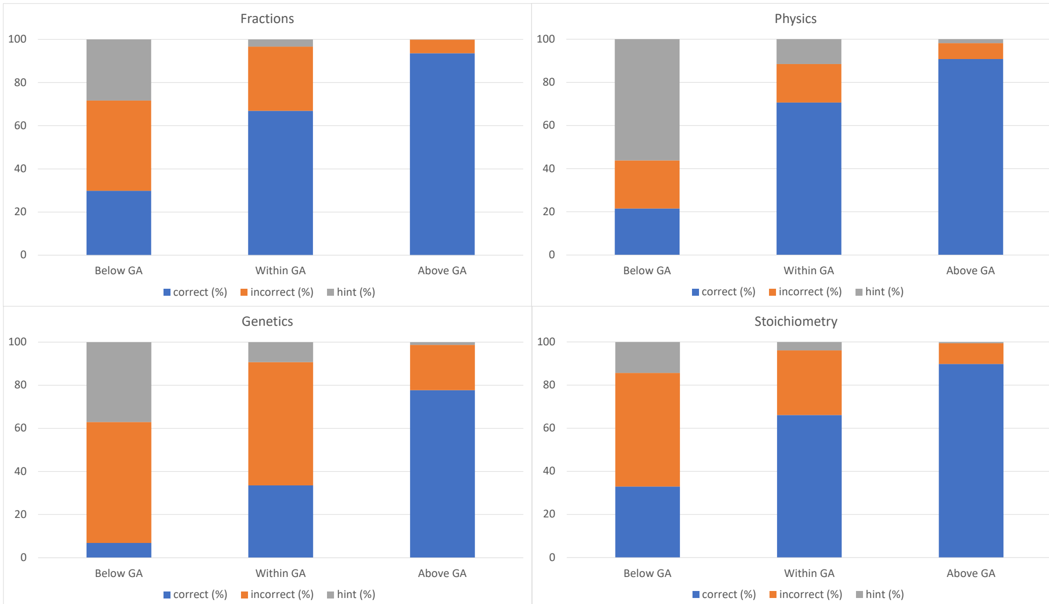

Fig. 2. Percentage ratio of correct, incorrect and hints.

	Below GA			Within GA			Above GA		
	h-c	h-h	h-i	h-c	h-h	h-i	h-c	h-h	h-i
Fractions	5	77	8	70	81	27	2	0	0
Genetics	15	431	74	46	147	51	3	0	0
Physics	626	2618	411	4678	3033	1407	375	34	38
Stoichiometry	210	972	307	983	510	509	87	20	8

Table 3. The distribution of help-seeking sequences (hint-correct, hint-hint and hint-incorrect) within the Grey area and its adjacent areas, below and above.

Next, we calculated the number of incorrect answers above and below the GA for each student in the four datasets to answer **RQ2**. Consequently, we compared the two groups using the Wilcoxon Signed Rank Test for paired samples as earlier. The results are presented in Table 5. In three out of four datasets (Fractions, Genetics, and Stoichiometry), the students provided, on average, more incorrect answers when below the GA than above. However, only for one case (Genetics) this difference was statistically significant. On the contrary, for the Physics dataset, the students provided more incorrect answers above the GA than below.

For **RQ3**, we compared the number of hint-correct sequences per student within and below the GA. According to the ZPD, a student who receives help when they are within their ZPD will be able to complete the learning task successfully. Therefore, if the GA is an effective proxy of the ZPD, then the

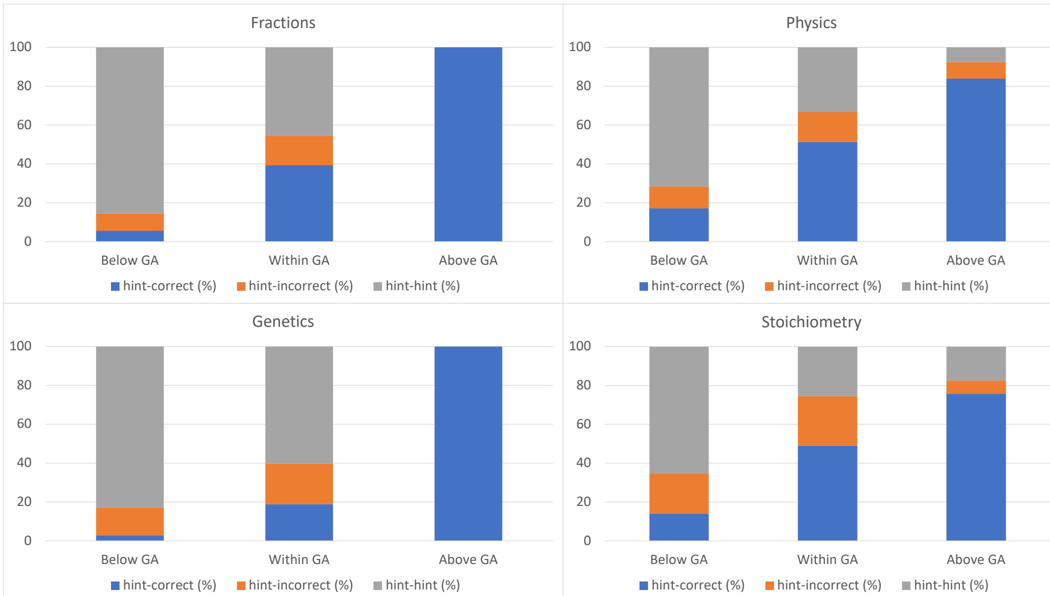

Fig. 3. Percentage ratio of help-seeking sequences: hint-correct (h-c), hint-hint (h-h) and hint-incorrect (h-i.

hint-correct sequences within the GA should be more than below the GA. The results are presented in Table 6. Indeed, the students on average perform more hint-correct sequences when they are within the GA than below. This difference is statistically significant for three out of four datasets (Fractions, Physics, Stoichiometry). For the Genetics dataset, although we established a difference, the result is not statistically significant.

5 Discussion

5.1 Validation of the GA approach over different ITSs and domains

To validate the effectiveness of the GA approach, we explored the following research questions: a) there would be more correct answers above the GA than below the GA (**RQ1**); b) there would be more incorrect answers below the GA than above the GA (**RQ2**); and c) There would be more hint-correct response sequences within the GA than below (**RQ3**) thus suggesting that students within the GA can carry out the learning task successfully when given appropriate support. A comparison of the frequencies of students' answers within the different areas confirmed the assumption that there are more correct answers above the GA than below the GA (RQ1). For all datasets, between 78% to 94% of the answers above the GA are correct while 7% to 33% of the answers are correct below the GA. In addition to these findings, the results of paired sample Wilcoxon

	$\text{correct}_{aboveGA}$ Mean	SD	$\text{correct}_{belowGA}$ Mean	SD	Wilcoxon Paired Samples V
Fractions	19	13	1.47	1.98	2883^{***}
Genetics	8.41	11.91	2.73	2.76	872
Physics	363.86	329.98	26.78	58.55	2115^{***}
Stoichiometry	62.28	89.33	3.52	7.05	48065^{***}

Table 4. The mean and standard deviation of correct answers above and below the GA per student for all four datasets. Bonferroni correction has been applied to account for multiple comparisons.

	$\text{incorrect}_{aboveGA}$ Mean	SD	$\text{incorrect}_{belowGA}$ Mean	SD	Wilcoxon Paired Samples V
Fractions	1.27	1.12	2.05	5.73	1304
Genetics	2.27	4.47	22.25	13	52^{***}
Physics	29.35	32.8	27.71	31.14	1043
Stoichiometry	5.17	6.95	6.63	11.20	29011

Table 5. The mean and standard deviation of incorrect answers above and below the GA per student for all four datasets. Bonferroni correction has been applied to account for multiple comparisons.

Signed Rank Tests show that students, on average, provide more correct answers above the GA than below. This difference was statistically significant for three out of four datasets (Fractions, Physics and Stoichiometry). Students interacting with these ITSs were significantly more likely to answer a question correctly if it was above the respective GA than below the GA.

Similarly, students provided more incorrect answers below the GA than above the GA (RQ2). However, for individual students' performance, this difference was statistically significant only in the case of the Genetics dataset. Our findings suggest that the number of correct and incorrect answers above and below the GA relates to the number of hints students request from the tutor and to the task's difficulty, as depicted by the overall amount of correct and incorrect answers. For example, the Genetics dataset had 54% incorrect first attempts as opposed to 29% of correct attempts and 17% hints. The other three datasets demonstrated over 70% correct responses. This may indicate that the content of the Genetics tutor was more difficult than the rest impacting on the distribution of correct and incorrect answers in the areas below and above the GA.

Finally, students generally used more hints below and within the GA than above. Students solved more steps correctly within the GA than they did below the GA. Above the GA, nearly no hints were used, but most answers following hints were correct. These results align with the concept of the ZPD. For all datasets, students on average, provided more correct responses after receiving a hint when they were within the GA than below (RQ3).

	$HC_{withinGA}$ Mean	$HC_{withinGA}$ SD	$HC_{belowGA}$ Mean	$HC_{belowGA}$ SD	Wilcoxon Paired Samples V
Fractions	0.91	1.42	0.06	0.29	553^{***}
Genetics	0.90	1.5	0.29	0.94	182
Physics	70.88	61.4	9.48	20.79	2021^{***}
Stoichiometry	2.60	3.98	0.56	2.12	22712^{***}

Table 6. The mean and standard deviation of hint-correct response sequences within and below the GA per student for all four datasets. Bonferroni correction has been applied to account for multiple comparisons.

Overall, less than 17% of the steps corresponded to hints (Genetics) while in two cases (Fractions and Stoichiometry) hints accounted for only 4% of the steps. This is consistent with the findings of previous studies about the usage of hints in ITSs. [18] examined the usage of hints within the Physics tutoring system Andes, and discovered that students did not use hints often enough when interacting with this tutor. This may point toward hint refusal, that is students rather guessing the answer instead of using hints [24, 21]. [1] explored the students' usage of hints in a Geometry tutor and reported that students were about ten times more likely to edit a wrong explanation than to ask for a hint. The authors conclude that giving learners total control over when they get hints may not be the best solution. Giving unsolicited hints instead might improve students' use of ITSs [24].

Our findings showed that below the GA most hints were followed by an additional hint request (h-h) rather than an incorrect answer (h-i), as one may have expected. [2] showed that an insufficient amount of information in a hint can cause frustration and the desire to request subsequent hints without attempts to solve a problem. Therefore, hint-hint sequences could be an additional indication of students' not being able to use the hints meaningfully, as a consequence of not being in their ZPD.

5.2 Theoretical and Practical Implications

The GA could potentially be an effective proxy of the ZPD based on the proposed metric of help-seeking effectiveness. However, we acknowledge that there are several factors that could influence the GA approach. First, the approach assumes that a model's uncertainty indicates the ZPD of a student. This is applicable in practice if the student model accurately predicts a student's performance. A well-performing student model and the choice of an optimal classification threshold are crucial for the effectiveness of the GA approach. In order to account for this when using the GA approach, the student model could be trained by using existing data of ITSs that are similar to the used ITS (e.g., similar learning activities). This could help to improve the predictive accuracy and find an optimum classification threshold, which is the basis for the definition of the GA.

One should keep in mind that the students' actual performance might be dependent on several factors which are not depicted by the model. For example, the optimal learning conditions differ for each learner and for different contexts [20]. Consequently, the size and shape of the ZPD might vary for each student and for each learning environment. Here, we used a GA of fixed size per dataset, thus assuming that students have similar ZPD. Nonetheless, ZPD can potentially be dependent on students' characteristics, such as age, or motivation. Furthermore, the effectiveness of the GA approach depends on the assistance provided by the ITS. Students within the GA might actually be able to solve a problem with appropriate assistance (i.e. student is within the ZPD), but if the provided assistance is not appropriate, the student may not succeed. At the same time, hints can only be helpful if they are actually requested. Our results are consistent with prior findings and indicate that students may refrain from using hints [1, 18, 24]. We argue that the GA can be used to address hint refusal by providing automatic hints to students who are predicted to be within the GA [10, 19].

6 Conclusion

This work investigated the effectiveness of the GA approach, a computational method for modeling the ZPD using metrics that relate help-seeking behavior to students' performance. We analyzed four datasets from students' interactions with different ITSs. For three out of the four datasets, the results confirmed that students were more likely to answer a question correctly if this question fell in the area above the GA than if it was below the GA. Furthermore, if students needed a hint for answering a question within the GA, the subsequent answer was more often correct than if a hint was used in order to answer a question below the GA. However, all three datasets consisted overall of more correct answers than incorrect. In the case the dataset consisted of more incorrect answers than correct, the opposite assumption was confirmed; that is, students were more likely to provide more incorrect answers below the GA than above. This may suggest that the task difficulty influences the effectiveness of the GA as a proxy of the ZPD, along with the students' attitude toward hints.

One limitation of this work relates to using existing data from an online repository and thus, having limited contextual information about study conditions and data collection processes. It is important to note that the present analysis does not include data associated with combined KCs. This simplification and the rare usage of hints resulted in small sample sizes potentially reducing the explanatory power of the tests. Finally, the present analysis was limited to the performance of students and their usage of hints within the three different areas above, within, and below GA. From this analysis, assumptions about other factors related to the ZPD (for example, size and symmetry of ZPD, influencing factors like the motivation of students) can be made, but these conclusions are not verified and should be further explored in future research.

Acknowledgements. We used the datasets 'Classroom study June 2013', 'MTLSD REAL GI GR Study 2015', 'USNA Physics Fall 2006', 'StoichStudy 4b Modality Effect' accessed via DataShop (pslcdatashop.web.cmu.edu)

References

1. Aleven, V., Koedinger, K.R.: Limitations of student control: Do students know when they need help? In: International conference on intelligent tutoring systems. pp. 292–303. Springer (2000)
2. Anohina, A.: Advances in intelligent tutoring systems: Problem-solving modes and model of hints. International Journal of Computers Communications & Control **2**(1), 48–55 (2014). https://doi.org/10.15837/ijccc.2007.1.2336
3. Cen, H., Koedinger, K., Junker, B.: Comparing two irt models for conjunctive skills. In: Intelligent Tutoring Systems: 9th International Conference, ITS 2008, Montreal, Canada, June 23-27, 2008 Proceedings 9. pp. 796–798. Springer (2008)
4. Chaiklin, S.: The zone of proximal development in vygotsky's analysis of learning and instruction. In: Kozulin, A., Gindis, B., Ageyev, V.S., Miller, S.M., Vygotskij, L.S. (eds.) Vygotsky's educational theory in cultural context, pp. 39–64. Learning in doing, Cambridge Univ. Press, Cambridge (2005). https://doi.org/10.1017/CBO9780511840975. 004
5. Chounta, I.A., Albacete, P., Jordan, P., Katz, S., McLaren, B.M.: The grey area: A computational approach to model the zone of proximal development. In: Lavoué, É., Drachsler, H., Verbert, K., Broisin, J., Pérez-Sanagustín, M. (eds.) Data driven approaches in digital education, Lecture notes in computer science Information systems and applications, incl. internet/web, and HCI, vol. 10474, pp. 3–16. Springer, Cham (2017)
6. Corbett, A., Koedinger, K., Anderson, J.: Intelligent Tutoring Systems, p. 849–874. Elsevier (1997). https://doi.org/10.1016/b978-044481862-1/50103-5, http://dx.doi.org/10.1016/b978-044481862-1/50103-5
7. Corbett, A., Kauffman, L., Maclaren, B., Wagner, A., Jones, E.: A cognitive tutor for genetics problem solving: Learning gains and student modeling. Journal of Educational Computing Research **42**(2), 219–239 (2010). https://doi.org/10.2190/ec.42.2.e
8. Desmarais, M.C., Baker, R.S.J.d.: A review of recent advances in learner and skill modeling in intelligent learning environments. User Modeling and User-Adapted Interaction **22**(1-2), 9–38 (2012). https://doi.org/10.1007/s11257-011-9106-8
9. D'Mello, S.K., Graesser, A.: Intelligent Tutoring Systems, p. 603–629. Routledge (Sep 2023). https://doi.org/10.4324/9780429433726-31, http://dx.doi.org/10.4324/9780429433726-31
10. Fossati, D., Di Eugenio, B., Ohlsson, S., Brown, C., Chen, L.: Generating Proactive Feedback to Help Students Stay on Track, p. 315–317. Springer Berlin Heidelberg (2010). https://doi.org/10.1007/9783642134371_56, http://dx.doi.org/10.1007/9783642134371_56
11. Katz, S., Albacete, P., Chounta, I.A., Jordan, P., McLaren, B.M., Zapata-Rivera, D.: Linking dialogue with student modelling to create an adaptive tutoring system for conceptual physics. International Journal of Artificial Intelligence in Education **31**(3), 397–445 (2021). https://doi.org/10.1007/s40593-020-00226-y
12. Katz, S., Connelly, J., Wilson, C.: Out of the lab and into the classroom: An evaluation of reflective dialogue in andes. Frontiers in artificial intelligence and applications **158**, 425 (2007)

13. Koedinger, K., Pavlik Jr, P.I., Stamper, J., Nixon, T., Ritter, S.: Avoiding problem selection thrashing with conjunctive knowledge tracing. In: Educational data mining 2011 (2010)
14. Koedinger, K.R., Baker, R.S., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J.: A data repository for the edm community: The pslc datashop. Handbook of educational data mining **43**, 43–56 (2010)
15. Luckin, R., Du Boulay, B.: Reflections on the ecolab and the zone of proximal development. International Journal of Artificial Intelligence in Education **26**(1), 416–430 (2016). https://doi.org/10.1007/s40593-015-0072-x
16. Margolis, A.A.: Zone of proximal development, scaffolding and teaching practice. Cultural-Historical Psychology **16**(3), 15–26 (2020). https://doi.org/10.17759/chp.2020160303
17. McLaren, B.M., DeLeeuw, K.E., Mayer, R.E.: Polite web-based intelligent tutors: Can they improve learning in classrooms? Computers & Education **56**(3), 574–584 (2011). https://doi.org/10.1016/j.compedu.2010.09.019
18. Muldner, K., Burleson, W., van de Sande, B., VanLehn, K.: An analysis of students' gaming behaviors in an intelligent tutoring system: predictors and impacts. User Modeling and User-Adapted Interaction **21**(1), 99–135 (2011). https://doi.org/10.1007/s11257-010-9086-0
19. Murray, R.C., VanLehn, K.: Effects of dissuading unnecessary help requests while providing proactive help. In: AIED. pp. 887–889 (2005)
20. Murray, T., Arroyo, I.: Toward measuring and maintaining the zone of proximal development in adaptive instructional systems. In: Cerri, S., Gouardères, G., Paraguaçu, F. (eds.) Intelligent Tutoring Systems. pp. 749–758. Springer (2002)
21. Ranganathan, R., VanLehn, K., van de Sande, B.: What do students do when using a step-based tutoring system? Research and Practice in Technology Enhanced Learning **9**(2), 323–347 (2014), https://rptel.apsce.net/index.php/rptel/article/view/2014-09019
22. Rau, M.A., Aleven, V., Rummel, N., Pardos, Z.: How should intelligent tutoring systems sequence multiple graphical representations of fractions? a multi-methods study. International Journal of Artificial Intelligence in Education **24**(2), 125–161 (2014). https://doi.org/10.1007/s40593-013-0011-7
23. Shabani, K., Khatib, M., Ebadi, S.: Vygotsky's zone of proximal development: Instructional implications and teachers' professional development. English Language Teaching **3**(4), 237–248 (2010), https://eric.ed.gov/?id=ej1081990
24. VanLehn, K.: Reflections on andes' goal-free user interface. International Journal of Artificial Intelligence in Education **26**(1), 82–90 (2016). https://doi.org/10.1007/s40593-015-0067-7
25. Vygotsky, L.: Interaction between learning and development. In: Gauvain, Cole (eds.) Readings on the Development of Children, pp. 34–40. Scientific American Books, New York (1978)