The “Grey Area”: A computational approach to model the Zone of Proximal Development

Irene-Angelica Chounta[1], Patricia Albacete[2], Pamela Jordan[2], Sandra Katz[2], Bruce M. McLaren[1]

[1] Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh PA, USA
[2] Learning Research and Development Center, University of Pittsburgh, Pittsburgh PA, USA
{ichounta, bmclaren}@cs.cmu.edu
{palbacet, pjordan, katz}@pitt.edu

Abstract. In this paper, we propose a computational approach to model the Zone of Proximal Development (ZPD) using predicted probabilities of correctness while students engage in reflective dialogue. We employ a predictive model that uses a linear function of a variety of parameters, including difficulty and student knowledge, as students use a natural-language tutoring system that presents conceptual reflection questions after they solve high-school physics problems. In order to operationalize our approach, we introduce the concept of the “Grey Area”, that is, the area of uncertainty in which the student model cannot predict with acceptable accuracy whether a student is able to give a correct answer without support. We further discuss the impact of our approach on student modeling, the limitations of this work and future work in systematically and rigorously evaluating the approach.

Keywords: Natural-language tutoring systems, intelligent tutoring systems, student modeling, zone of proximal development.

1 INTRODUCTION

Intelligent Tutoring Systems (ITSs) support students in grasping concepts, applying them during problem-solving activities, addressing misconceptions and in general improving students’ proficiency in science, math, reading and other areas [1]. However, we still face the challenge of developing tutoring systems that emulate the interactive nature of human tutoring and that are just as effective – if not better – than human tutors. One approach for achieving this goal is to engage students in reflective discussions about scientific concepts [2]. To a large extent, these systems lack the ability to gauge students’ level of mastery over the domain content that the tutoring system was designed to support. This is also challenging for human tutors, who roughly assess the level of knowledge and understanding of their tutees, although they are generally poor at diagnosing the specific causes of student errors [3]. We argue that a student model that maintains and dynamically updates a representation of students’ ability level on targeted curriculum elements can help us bridge the gap between simulated and human tutors. Tutorial dialogue systems could be more effective if they are guided by the information about the student's understanding of curriculum elements that is represented within a student model, along with other student characteristics such as demographic information and motivational factors such as interest in the targeted domain, self-efficacy, etc. [4].

1.1 Research Hypothesis and Impact

In order to provide meaningful instruction and scaffolding to students, a tutoring system should appropriately adapt the learning material with respect to both content and presentation. A way to achieve this is to dynamically assess the student's knowledge state and needs. Human tutors use their assessment of student ability to adapt the level of discussion to the student's "zone of proximal development" (ZPD) [5]. Adapting the conversation to the ZPD would mean asking the student questions just beyond their knowledge level – in other words, asking questions that students are able to answer correctly with adequate support. During tutorial dialogues, human teachers evaluate their students' learning state; a teacher judges whether a student will be able to answer a question correctly without any help (that is, the student is above the ZPD) or be able to answer correctly if given some help (that is, the student is in the ZPD) or unable to answer correctly even with help (that is, the student is below the ZPD). Depending on this judgment, the teacher will choose to ask this question, provide hints, or instead choose a more appropriate question for this student's ability level. We propose a computational approach for modeling students' ZPD as they carry out learning activities using a dialogue-based intelligent tutoring system, which replicates the pedagogical strategies of human tutors. The predictions of the student model serve as a proxy for human tutors' judgment. In particular, we employ a student model to assess students' changing knowledge state as they engage in a dialogue with the system. At each step of the dialogue, the student model predicts the probability of the student being able to answer the question posed by the computer tutor correctly. When the predicted probability is high, the student is likely to possess the knowledge needed to answer the question correctly. When the predicted probability is low, it is unlikely that the student has an adequate grasp of the necessary knowledge to give a correct response. An interesting case arises when the student model predicts that the student will be able to answer a question with a probability around 50%, because in this case there is greater uncertainty. In other words, the student may need some extra support to be able to give a correct response. Hence the region of predicted probabilities that reflects this area of uncertainly with regards to the student's abilities to give a correct answer without support is what we call the "Grey Area" [6]. Our research hypothesis is that we can use the fitted probabilities as predicted by the student model to model the ZPD. The core rationale is that if the student model cannot predict with acceptable accuracy whether a student will answer a question correctly, then it might be the case that the student is in the ZPD. To the best of our knowledge, this is a novel approach to modeling the ZPD, never before implemented or reported in the literature.

In the following section we discuss relevant research about the ZPD, Intelligent Tutoring Systems and student modeling. In section 3, we present our approach and methodology. Analysis and results are presented in section 4. Finally, in section 5 we discuss the impact and implications of our approach and conclude by presenting the limitations of our study and future work.

2 RELATED WORK

2.1 Zone of Proximal Development

The Zone of Proximal Development (ZPD) is one of the best known concepts in educational psychology, defined by Vygotsky as: "the distance between the actual developmental level as determined by independent problem-solving and the level of potential development as determined through problem-solving under adult guidance or in collaboration with more capable peers" [5]. This definition of the ZPD indicates the importance of appropriate assistance in relation to the learning and development process and thus it can be stated more simply as "the difference between what a learner can do without help and what he or she can do with help" [7]. Deriving ways to identify and formally describe the ZPD is an important step towards understanding the mechanisms that drive learning and development, gaining insights about learners' needs, and providing appropriate pedagogical interventions [8]. Approaches to identifying and/or modeling the ZPD typically depend on finding instances of successful assisted performance; for example, tasks that a student carries out successfully after having received some kind of scaffolding [8]. Various methods that derive from or build upon this notion have been developed for the dynamic assessment (DA) of the learning potential of students (or learners in general) [9]. Usually these approaches employ tests that measure the difference between unmediated and mediated performance [10] or the cognitive modifiability of learners (i.e., how students' cognitive structures change when they fail a task and the teacher/expert gives them help or remediation tasks) [11]. However, Dynamic Assessment focuses on assessing the learning or development potential of the learner rather than the actual level of development. Luckin and du Boulay proposed the use of domain knowledge representations and Bayesian Belief Networks (BBN) to construct the Zone of Proximal Adjustment (ZPA) [12], that is the tutor's adaptation mechanism to the ZPD of particular learners. Each student's knowledge is represented as an overlay model and the student model is compared to the domain knowledge representation.

2.2 Intelligent Tutoring Systems and Student Modeling

Intelligent Tutoring Systems (ITSs) commonly use student models to track the performance of students and choose appropriate content for practicing skills and fostering knowledge. Most student models developed for ITSs are based on the notion of mastery learning; that is, the student is asked to continue solving problems or answering questions about a concept until she has mastered it. Only then will the student be guided to move forward to other concepts [13]. Mastery learning is in line with the notion of learning curves [14] that is, how many opportunities a student needs to master a skill. One could argue that mastery learning is consistent with the ZPD, in the sense that the student is considered to have mastered a skill when she is able to successfully carry out a task that requires this particular skill without help. However, the ZPD does not directly address mastery but rather potential "development" under appropriate assistance; by identifying the ZPD not only can we assess the state of a student's knowledge but we also gain insight into how appropriate instruction can scaffold development [15]. Human tutors do not carry out detailed diagnoses of student knowledge and their assessments of students' knowledge are often inaccurate [3]. Nevertheless, they typically construct and dynamically update a normative mental representation of students' understanding, as reflected in tutors' adaptive responses to students' need for scaffolding or remediation [16]. Similarly, a tutorial dialogue system uses a student model to adapt to the student's needs. Otherwise, all students would be presented with the same topics, at the same level of detail or complexity. Moreover, if the student answers a question incorrectly and there is need for remediation, the simulated tutor will not be able to adapt the type of support that it provides. Indeed, it is the absence of information about the student that forces designers of tutorial dialogue systems to make a "best guess" about how to structure a dialogue—that is, what the main "line of reasoning" should be, what remedial or supplemental sub-dialogues to issue and when—and then to hard code these guesses into the dialogues. Consequently, with the "one size fits all" approach to dialogue that is implemented in most tutorial dialogue systems, students are often under-exposed to material that they don't understand and overexposed to material they have a firm grasp of. The first problem renders these systems ineffective in enabling students to achieve mastery over the focal content; the second makes them inefficient. Developing a computational model of students' ZPD takes an important step towards generating more adaptive tutorial dialogues.

3 METHODS

3.1 Rimac: A Dialogue Tutor for Physics

In this study we explored the proposed approach using Rimac, a web-based natural-language tutoring system that engages students in reflective discussions about concepts after they solve quantitative physics problems [17]. Rimac has been used successfully to teach physics concepts to high-school students. We used data collected during three previous studies with Rimac to train a student model and predict students' performance. The three studies were conducted within high school physics classes at schools in the Pittsburgh, PA (U.S.) area, following a similar protocol. First, students took a pretest and were introduced to Rimac. Then they interacted with Rimac to discuss the physics knowledge associated with quantitative problems on dynamics. Finally, students took a post-test to measure learning gains. The tests aimed to test students' conceptual understanding of physics instead of their ability to solve quantitative problems. Rimac's dialogues were developed to present a directed line of reasoning, or DLR [18], in which the tutor presents a series of questions to the student. If the student answers a question correctly, she advances to the next question in the DLR. If the student responds incorrectly, the system launches a remedial sub-dialogue and then returns to the main line of reasoning after the sub-dialogue has completed. If the system is unable to understand the student's response, it completes the step for the student (for more details, see [19]). The knowledge components related to tutor question/student response pairings are logged during the system's interactions with students and were used to train the student model as described next. A short example of a dialogue with Rimac is presented in **Table 1**.

Table 1. A short example of an adaptive dialogue with Rimac

Tutor:	So, can you please tell me what the vertical forces on the arrow are?
Student:	gravity
Tutor:	Very good. Since we know that the force of gravity is acting on the arrow, what does that mean about the arrow's vertical acceleration (zero, nonzero, etc)?
Student:	nonzero

3.2 The Student Model

For this study, we used an Additive Factor Model (AFM), introduced by Cen et. al. [20], to model students' knowledge. The model uses logistic regression to predict the probability of a student *i* completing a step *j* correctly as a linear function of student parameters (the student's proficiency θ_i), skill parameters β_k and the learning rates of skills γ_k, as shown in equation (1). AFM takes into account the frequency of prior practice and exposure to skills but not the correctness of responses since it assumes all students accumulate knowledge in the same way. In this paper we implemented the AFM model following the approach of Chi et al. [21] who modeled students working on physics problems using a dialogue-based tutor.

$$ln \frac{p_{ij}}{1 - p_{ij}} = \theta_i + \sum_k \beta_k Q_{kj} + \sum_k Q_{kj}(\gamma_k N_{ik}) \quad (1)$$

The dataset was collected by training 291 students on Rimac over a period of 4 years (2011-2015). During students' interactions with Rimac, they answered reflection questions on physics problems about dynamics, such as: "*In our first question we will focus on the horizontal motion of the arrow. Let`s imagine a scenario in which an archer is standing at the edge of a high cliff. He shoots an arrow perfectly horizontally with an initial velocity of 60 m/s off this cliff. During the arrow`s flight, how does its horizontal velocity change (increases, decreases, remains the same, etc.)? Remember that you can ignore air resistance*". Students worked on reflection questions about three physics problems that explored motion laws and addressed 88 knowledge components (KCs). The dataset contained in total 15,644 student responses. Each student response answers a question posed by the tutor and was classified as correct or incorrect. For the training of the model we split our dataset following an 80-20 rule [22]: 12,515 student responses were used for training the model and the remaining 3,129 were used for testing. On average, each student answered a total of 53 questions, stemming from several reflection questions. The test set contained on average 11 entries per student.

3.3 The Grey Area and the Study Setup

To predict the correctness of students' responses, we used the aforementioned AFM student model. Then, we classified the outcome as correct or incorrect based on the fitted probabilities provided by the model. In this study, the student model provided predictions at the step level (one step is one question/answer of the tutorial dialogue). A step might involve one or multiple KCs. The classification threshold in this case (i.e., the cutoff determining whether a response is classified as correct or incorrect) is 0.5 and was validated using the receiver operating characteristic (ROC) curve for the binary classifier (**Fig. 1**). For example, if the fitted probability for a step in the dialogue is 0.8 (above 0.5) then we expect that the student will be able to answer the corresponding dialogue step correctly; hence, it is classified as correct. Similarly, if the fitted probability for a step in the dialogue is 0.2 (below 0.5) then we expect that the student will not be able to answer correctly; hence it is classified as incorrect.

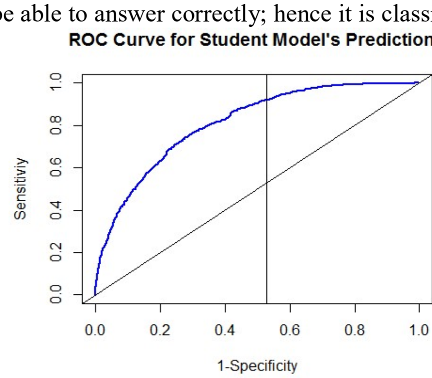

Fig. 1. ROC Curve for validating the classification threshold

We found that the predicted probabilities correlate with students' performance in the pre and post-tests: the student model will provide high probabilities of correctness (i.e., a high probability to answer a question correctly) for students who performed well on the pre and post-tests. Similarly, the model will provide low probabilities of correctness for students who performed poorly in the pre and post-tests. We argue that this correlation between students' performance in the pre and post-tests and predicted probabilities suggests that the predicted probabilities are appropriate indicators of the ZPD. In this study the pre and post-tests assessed conceptual knowledge associated with the questions that students were assigned to work on. Furthermore, we expect that the closer the prediction is to the classification threshold, the higher the uncertainty of the model and thus, the higher the prediction error. In other words, when the student model predicts that the student will be able to answer a question with a probability close to 0.5, we are more uncertain than with any other prediction as to whether or not the student will answer the question correctly. According to our hypothesis, the window where the prediction error is high (i.e. the "Grey Area") can be used to approximate the student's zone of proximal development. The concept of the Grey Area is depicted in **Fig. 2**.

The space “Above the Grey Area” denotes the region where the student is predicted to answer correctly (the fitted probability is considerably higher than the cutoff threshold) and consequently may indicate the area above the ZPD; that is, the area in which the student is able to answer a question without any assistance. Accordingly, the space “Below the Grey Area” denotes the area where the student is predicted to answer incorrectly (the fitted probability is considerably lower than the cutoff threshold) and consequently may indicate the area below the ZPD; that is, the area in which the student is not able to carry out the task either with or without assistance. In this paper, we model the Grey Area symmetrically around the classification threshold for simplicity and because the binary classifier was set to 0.5. However, the symmetry of the Grey Area is something that could change depending on the classification threshold and the learning objectives. This is also the case for the size of the Grey Area.

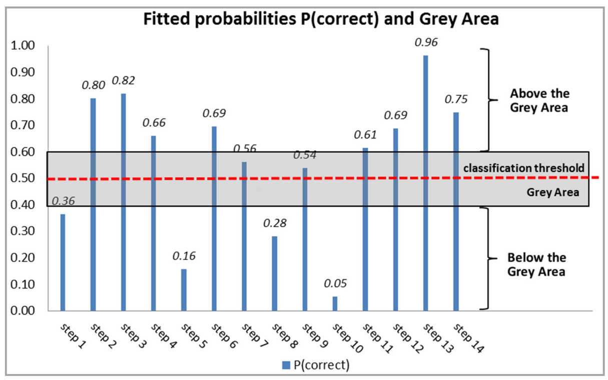

Fig. 2. The Grey Area concept with respect to the fitted probabilities as predicted by the student model for a random student and for the various steps of a learning activity. Here we depict the Grey Area ranging from 0.4 to 0.6 and extending on both sides of the classification threshold (dotted line).

In this paper we present the concept of the Grey Area and the methodology to model the ZPD. We are exploring, but have not yet specified several design aspects (e.g. thresholds, the use of symmetrical or asymmetrical Grey Areas etc.). We do not propose a specific size but rather experiment with Grey Areas of different sizes and study how the student model behaves within these areas. We believe that the decision about the appropriate size (or shape) of the Grey Area is not only a modeling issue but also, and perhaps predominantly, a pedagogical one since it relies on the importance of the concepts taught, the teaching strategy and the learning objectives. That is, a teacher may consider it important to elicit an answer for a question even if it is predicted that the student is unable to correctly answer that question. The student may not be knowledgeable enough about the given topic or it might be of minor importance and thus even a low probability of correctness would be considered sufficient to classify the student as knowledgeable.

4 ANALYSIS AND RESULTS

4.1 Model Behavior and Student Performance

Our research hypothesis is rooted in the belief that the predicted probabilities of the student model can provide insight into student knowledge and performance. That is, the fitted probabilities for a high-performing student will be higher than the fitted probabilities for a low-performing student. One could argue that predicted probabilities are a model's characteristic and may not be appropriate to describe students' performance. However, the fitted probability represents the probability that a student will correctly answer a dialogue question. Since high performers have a higher probability of correctly answering questions, the average of their fitted probabilities will be higher than those of low performing students.

We performed a correlation analysis to explore this hypothesis. We correlated the average fitted probability (i.e., the average value of the fitted probabilities for the answers of each student) per user with the students' knowledge pre-test scores. The correlation analysis showed that the average fitted probability correlates positively with the pre-test scores at a statistically significant level (Pearson's $r = 0.396^{**}$, $p<0.01$). The positive correlation was also confirmed for the post-test scores (Pearson's $r = 0.46^{**}$, $p<0.01$). This suggests that if a student scores high on the pre-test for a particular KC, the model will predict that this student is able to answer a question that deals with this KC. Similarly, a student who was predicted to answer correctly a question dealing with a KC will also have a high post-test score for this KC. This finding indicates that the model can predict a student's performance and may be further used to model the student's zone of proximal development. One might notice that the correlation between the average fitted probabilities and the pre and post-knowledge tests are not high (Pearson's $r < 0.5$). However, this might be due to the fact that in the pre and post knowledge tests we only test a small number of the knowledge components that are present in the dialogues. Therefore, the pre and post-knowledge scores can be suggestive of the student's knowledge state but they do not accurately represent it. Model Accuracy for cases inside the Grey Area

The Grey Area is defined as the area where the model cannot accurately predict whether a student will correctly answer a particular question. To operationalize the grey area with respect to size and threshold, we define areas of different sizes and further explore the model's behavior within these areas. For this study, we considered five grey areas of different sizes: Area 1 ($0.45 < p <0.55$), Area 2 ($0.4 < p <0.6$), Area 3 ($0.35 < p <0.65$), Area 4 ($0.3 < p <0.7$) and Area 5 ($0.25 < p <0.75$). We chose these particular areas for symmetry and also to cover the range around the classification threshold for which one would expect low predictive accuracy. For these areas, we calculated how many times the model predicted the student answer accurately, where accuracy is defined as the total number of times (a) the student answered correctly and the model also predicted the student would answer correctly and (b) the student answered incorrectly and the model also predicted the student would answer incorrectly, divided by the total number of predictions. **Table 2** presents the non-cumulative and the cumulative analysis of the data. For non-cumulative analysis, we mean the analysis of the cases that are contained only in the focal grey area under study (non-cumulative results) and exclude the cases that are also contained in preceding areas. For example, in Area 2 we examine 420 cases that are not contained in Area 1. The cumulative analysis presents the analysis of cases that are contained in the current area but can also be part of the preceding grey area (cumulative results). For example, Area 2 analyzes 814 cases out of which 394 are also contained in Area 1. The results of the non-cumulative analysis show that most predicted cases fall in Area 2 – Non Cumulative (the largest increase in uncertain cases is with Area 2) and that 42.6% of them are predicted incorrectly. This means that for 13.4% (420 cases) of the total number of cases (Total Number of Cases: 3,129), the model gave a prediction with a probability from 0.4 to .45 and .55 to 0.6. As we move away from the classification threshold (0.5), the number of additional fitted cases tends to decrease (fewer cases are predicted with probabilities far from the cutoff threshold) but the percentage of the correct predictions improves. This is depicted in Fig. 3. That finding was expected since the confidence of the model increases.

Table 2. Predictions' accuracy within grey areas of different sizes.

NC / (C)	Area 1	Area 2	Area 3	Area 4	Area 5
#Cases - NC /(*C*)	394 /(*394*)	420 /(*814*)	404 /(*1218*)	369 /(*1587*)	304 /(*1891*)
Cases (%)- NC /(*C*)	12.6/(*12.6*)	13.4 /(*26.0*)	12.9 /(*38.9*)	11.8 /(*50.7*)	9.7 /(*60.4*)
#Correct- NC /(*C*)	213 /(*213*)	241 /(*454*)	259 /(*713*)	250 /(*963*)	221 /(*1184*)
#Incorrect- NC /(*C*)	181 /(*181*)	179 /(*360*)	145 /(*505*)	119 /(*624*)	83 /(*707*)
Correct (%)- NC /(*C*)	54.1 /(*54.1*)	57.4 /(*55.8*)	64.1 /(*58.5*)	67.8 /(*60.7*)	72.7 /(*62.6*)
Incorrect (%)- NC /(*C*)	45.9 /(*45.9*)	42.6 /(*44.2*)	35.9 /(*41.5*)	32.3 /(*39.3*)	27.3 /(*37.4*)

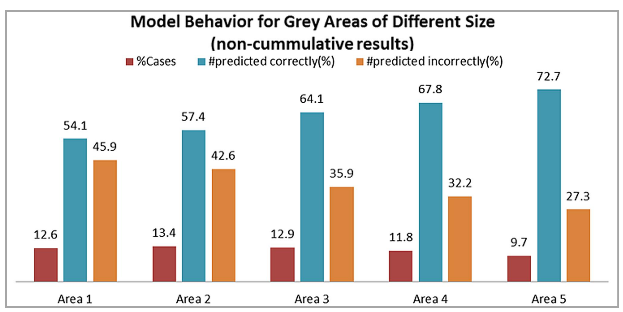

Fig. 3. Model behavior (percentage of total number of predicted cases, cases predicted correctly and cases predicted incorrectly) within the five grey areas of different sizes.

For Area 1, the prediction error is higher (45.9% of the cases were not predicted correctly) but the number of fitted cases is lower than Area 2. In Fig. 4 we depict the results for the cumulative analysis. As expected, more cases are predicted correctly as the size of the area increases. On one hand, choosing a narrow grey area to model the ZPD would limit the number of cases we scaffold since fewer cases would fall within the area. On the other hand, choosing a wide grey area would affect the accuracy;

that is, some cases that could be predicted correctly would be falsely labeled as "grey". Our work to date does not aim to define the appropriate size for the Grey Area but rather to study how the model's behavior changes for areas of different size. It is worth mentioning that for the area that is not included in the five areas we study—that is, the area [0,0.25)U(0.75, 1]—the model predicted 89% of the cases correctly while the overall accuracy of the model was 73%.

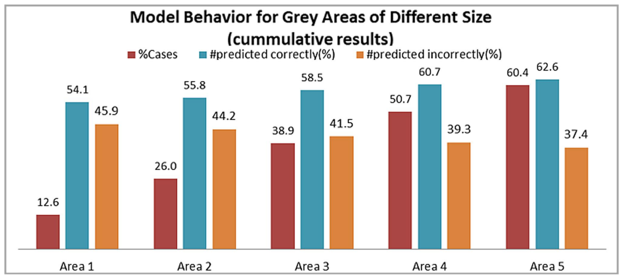

Fig. 4. Model behavior (percentage of total number of predicted cases, cases predicted correctly and cases predicted incorrectly) within the five grey areas of different sizes.

4.2 Grey Areas and Students' Performance

So far, we have studied how the model performs within grey areas of various sizes, but we have no indication of students' performance. One could argue that based on the way the grey zone was modeled—that is, symmetrical around the cutoff threshold of the binary classifier—correct and incorrect answers should be balanced and not vary significantly from one zone to the other. Again, here we only study students' performance; therefore "correctness" refers to the student's answers (i.e., whether a student answered a question correctly) and not whether the model predicted correctly (i.e., whether the model predicted that the student would answer the way she answered). For the five grey areas defined in 4.2, we have counted the number of correct and the number of incorrect student answers.

In **Fig. 5** we present the distribution of correct and incorrect answers over the different grey areas and over correct and incorrect model predictions (as shown in the cumulative analysis in **Fig. 4**). For example, for Area 1, the model predicted 54.1% of the cases correctly- that is, the model predicted that a student would answer correctly and indeed the student answered correctly, or the model predicted that a student would answer incorrectly and indeed the student answered incorrectly. Out of these cases, 28.7% were correct answers to the question involved and 25.4% were incorrect. Likewise, for Area 1 the model predicted 45.9% cases incorrectly. Out of these cases, 18% were correct answers to the question involved and 27.9% were incorrect. It is evident that even though the accuracy of the prediction changes between areas of different sizes, the distributions of correct and incorrect answers are similar. Another thing that can be noted is that for cases that the model predicts correctly, the ratio of correct/incorrect answers is around 1.2 (correct answers are slightly more than incorrect). On the contrary, for cases that are not predicted correctly by the model the ratio of correct/incorrect answers are about 0.7 signifying that incorrect answers outnumber correct ones. Nonetheless this is a pattern that is maintained for all of the grey areas and most probably it reveals that the student model tends to provide positive predictions.

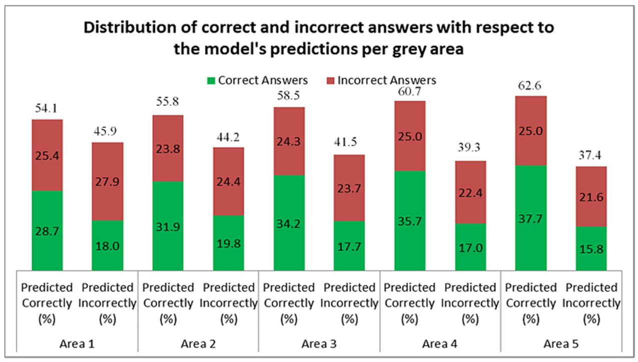

Fig. 5. Graphical representation of the distribution of correct and incorrect answers (percentage) with respect to the model's correct and incorrect predictions.

5 DISCUSSION

5.1 Contribution of the approach

We envision that the contribution of the proposed approach, besides its novelty, will be in defining and perhaps revising instructional methods to be implemented by ITSs. As noted previously, the most popular instructional method used to choose learning content (problems, activities, examples, etc.) is mastery learning. This means that the student goes through the same concept again and again until the probability of having mastered it is near certainty. Although mastery learning is highly effective—and might largely account for the effectiveness of human tutoring [1], it could lead to tedious repetition or frustration and eventually discourage the student from achieving the goal. Choosing the "next step" is a more prominent issue in the case of dialogue-based intelligent tutors. Not only should the task be appropriate with respect to the background knowledge of the student, but it should also be presented in an appropriate manner so that the student will not be overwhelmed and discouraged – if the task is hard for the student – or boring and not challenging – if the task is too easy. Another key distinction with Mastery Learning perhaps worth mentioning is the idea that ZPD focuses on the level of help. Mastery Learning implies that help might be needed to move the student forward, but it doesn't explicitly include it as part of the definition.

To address this issue, we need an assessment of the knowledge state of each student and insight into the appropriate level of support the student needs to achieve the learning goals. This is described by the notion of ZPD. We claim that our approach makes an explicit link between student modeling and the ZPD and that this approach is a reasonable and novel operationalization of the ZPD. It is evident that if we can model the ZPD then we can adapt our instructional strategy accordingly. For example, if a student is above the ZPD—that is, able to solve a problem on her own and without any help—the tutor will probably challenge the student with some questions that go beyond the current problem's level of difficulty. On the other hand, if a student is in the ZPD—that is, the student needs help and appropriate scaffolding to solve the current problem—the tutor will go slowly, perhaps clarifying step by step the knowledge the student seems to be lacking. Finally, if a student is below the ZPD—that is, the student completely lacks the necessary skills and will not be able to solve the problem, either with or without help—the tutor might choose to skip this problem or to select more appropriate (perhaps simpler) problems. Depending on the state of the student's knowledge, the tutorial dialogue may be directed and focus on particular curriculum elements (facts, concepts, skills, etc.) to discuss during a given problem and to determine the appropriate level at which to discuss these elements.

5.2 Validation of the proposed approach

In this paper, we provide preliminary support for our approach. It is also necessary to validate our approach. The challenge in doing so lies in the fact that there is no objective way to test that a student is (or is not) in the ZPD. One heuristic that could be used to explore this is to provide different levels of support to students using the proposed approach and then observe the outcome. Students who are expected to be in the ZPD and who receive appropriate scaffolding should be able to correctly answer the questions asked by the tutor. Thus, we plan to carry out larger scale studies where the dialogue will adapt to the student's knowledge according to the guidance provided by the student model and the represented Grey Area. The dialogue adaptation will take place on selected dialogue steps (in order to maintain the coherency of the dialogue) and will be implemented following three basic adaptation rules:

- Students who are above the Grey Area will receive more challenging questions, no help or even skip specific parts of the dialogue that the model predicts they have mastered;
- Students who are within the Grey Area will receive meaningful information, scaffolding and hints related to the step in question;
- Students who are below the Grey Area will either skip the step that the model predicts they are unable to answer or they will receive explicit information and instruction.

To evaluate our approach, we will study the learning gains of students who receive different levels of support (hints, worked out examples, explicit information, etc.) based on their performance in pre- and post- knowledge tests and their performance during activities within Rimac. We are optimistic that the dialogue adaptation according to the Grey Area concept will improve students' learning gains and motivation.

6 CONCLUSION

In this paper, we present a computational approach to model the Zone of Proximal Development in ITSs. To that end, we introduce the concept of the "Grey Area", that is the area of uncertainty in which the student model cannot predict with acceptable accuracy whether a student is able to give a correct answer without support. It is important to point out that we do not claim that the Grey Area is the ZPD. Instead, our proposal is that if the model cannot predict the state of a student's knowledge, it may be that the student's knowledge state falls withing the ZPD.

As an initial test to justify our hypothesis, we used data collected from classroom studies where students reflected on the concepts associated with physics problems, using a dialogue-based tutoring system (Rimac). We explored the operationalization of our approach by studying the behavior of the student model and the performance of students within grey areas of various sizes. We found that the accuracy of the model changes depending on the size of the grey zone but the distribution of correct and incorrect student responses remains fairly constant. Additionally, we showed that the average predicted probabilities per student—that is, the average value of the fitted probabilities for a particular student during her interaction with the Dialogue Tutor—correlates positively on a statistically significant level with the student's scores on pre- and post-knowledge tests. This suggests that the student model predictions can provide reliable indicators of students' performance. One limitation of our work is that we have not yet conducted a larger-scale and rigorous evaluation of the approach; however, plans to validate the model are being developed. Specifically, we plan to carry out extensive studies to explore the proposed approach to modeling the ZPD, as well as to better understand the strengths and limitations of using a student model to guide students through adaptive lines of reasoning.

Acknowledgements. This research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A150155 to the University of Pittsburgh. The opinions expressed are those of the authors and do not necessarily represent the views of the Institute or the U.S. Department of Education.

We thank Sarah Birmingham, Dennis Lusetich, and Scott Silliman for their contributions.

References

1. VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educ. Psychol. 46, 197–221 (2011).
2. Graesser, A.C., Person, N., Harter, D., Group, T.R., others: Teaching tactics and dialog in AutoTutor. Int. J. Artif. Intell. Educ. 12, 257–279 (2001).
3. Chi, M.T., Siler, S.A., Jeong, H.: Can tutors monitor students' understanding accurately? Cogn. Instr. 22, 363–387 (2004).
4. Chi, M., VanLehn, K., Litman, D., Jordan, P.: An evaluation of pedagogical tutorial tactics for a natural language tutoring system: A reinforcement learning approach. Int. J. Artif. Intell. Educ. 21, 83–113 (2011).

5. Vygotsky, L.: Interaction between learning and development. Read. Dev. Child. 23, 34–41 (1978).
6. Chounta, I.-A., McLaren, B.M., Albacete, P., Jordan, P., Katz, S.: Modeling the Zone of Proximal Development with a Computational Approach. In: Proceedings of the 10th International Conference on Educational Data Mining (EDM 2017).
7. Reber, A.S.: The Penguin dictionary of psychology. Penguin Press (1995).
8. Chaiklin, S.: The zone of proximal development in Vygotsky's analysis of learning and instruction. Vygotsky's Educ. Theory Cult. Context. 1, 39–64 (2003).
9. Tzuriel, D.: Dynamic assessment of young children: Educational and intervention perspectives. Educ. Psychol. Rev. 12, 385–435 (2000).
10. Poehner, M.E., Lantolf, J.P.: Bringing the ZPD into the equation: Capturing L2 development during Computerized Dynamic Assessment (C-DA). Lang. Teach. Res. 17, 323–342 (2013).
11. Feuerstein, R., Rand, Y., Jensen, M.R., Kaniel, S., Tzuriel, D.: Prerequisites for assessment of learning potential: The LPAD model. Dyn. Assess. 35–51 (1987).
12. Luckin, R., Du Boulay, B., others: Ecolab: The development and evaluation of a Vygotskian design framework. Int. J. Artif. Intell. Educ. 10, 198–220 (1999).
13. Corbett, A.T., Koedinger, K.R., Anderson, J.R.: Intelligent tutoring systems. Handb. Hum.-Comput. Interact. 5, 849–874 (1997).
14. Martin, B., Mitrovic, A., Koedinger, K.R., Mathan, S.: Evaluating and improving adaptive educational systems with learning curves. User Model. User-Adapt. Interact. 21, 249–283 (2011).
15. Hedegaard, M.: 10 The zone of proximal development as basis for instruction. Introd. Vygotsky. 227 (2005).
16. Katz, S., Allbritton, D., Connelly, J.: Going beyond the problem given: How human tutors use post-solution discussions to support transfer. Int. J. Artif. Intell. Educ. 13, 79–116 (2003).
17. Katz, S., Albacete, P.L.: A tutoring system that simulates the highly interactive nature of human tutoring. J. Educ. Psychol. 105, 1126 (2013).
18. Evens, M., Michael, J.: One-on-one tutoring by humans and computers. Psychology Press (2006).
19. Jordan, P., Albacete, P., Katz, S.: Exploring Contingent Step Decomposition in a Tutorial Dialogue System. In: The 24th Conference on User Modeling, Adaptation and Personalization (UMAP).
20. Cen, H., Koedinger, K., Junker, B.: Comparing two IRT models for conjunctive skills. In: International Conference on Intelligent Tutoring Systems. pp. 796–798. Springer (2008).
21. Chi, M., Koedinger, K.R., Gordon, G.J., Jordan, P., VanLehn, K.: Instructional factors analysis: A cognitive model for multiple instructional interventions. In: Proceedings of the 4th international conference on educational data mining. pp. 61–70 (2011).
22. Kiremire, A.R.: The application of the Pareto principle in software engineering. Consult. January. 13, 2016 (2011).