THE GREY AREA: TOWARDS A COMPUTATIONAL APPROACH FOR MODELING THE ZONE OF PROXIMAL DEVELOPMENT

Irene-Angelica Chounta, Bruce M. McLaren

Human Computer Interaction Institute, Carnegie Mellon University, USA

Patricia Albacete, Pamela Jordan, Sandra Katz

Learning Research and Development Center, University of Pittsburgh, USA

Paper presented at the 10th International Conference on Educational Data Mining

Wuhan, China, June 2017

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 are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

The Grey Area: Towards a Computational Approach for Modeling the Zone of Proximal Development

Irene-Angelica Chounta, Bruce M. McLaren
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh PA, 15213, USA
{ichounta,bmclaren}@cs.cmu.edu

Patricia Albacete, Pamela Jordan, Sandra Katz
Learning Research and Development Center
University of Pittsburgh
Pittsburgh PA, 15260, USA
{palbacet, pjordan, katz}@pitt.edu

ABSTRACT

In this paper, we propose a computational approach to modeling the Zone of Proximal Development of students who learn using a natural language tutoring system for physics. We employ a student model that predicts students' performance based on their prior knowledge and their activity when using a dialogue tutor to practice with conceptual, reflection questions about high-school physics. Furthermore, we introduce the concept of the "Grey Area", the area in which the student model cannot predict with acceptable accuracy whether a student has mastered the knowledge components or skills present in a particular step. We envision that our approach will contribute to the way we design learning content for ITSs and the way we author dialogues for natural-language tutoring systems. We further discuss the impact of our approach on student modeling and discuss 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 [25]. 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. ITS researchers and developers have been studying the use of simulated tutorial dialogues that aim to engage students in reflective discussions about scientific concepts [10, 13, 23]. However, to a large extent, these systems lack the ability to gauge students' level of mastery over the curriculum that the tutoring system was designed to support. This is also challenging for human tutors, who do gauge the level of knowledge and understanding of their tutees to some degree, although they are poor at diagnosing the causes of student errors [7].

We argue that integrating a student model into tutorial dialogue systems that maintain and dynamically update a representation of students' ability level on targeted curriculum elements can help us address these differences between human and simulated tutors. Correspondingly, we propose that tutorial dialogue systems would be more effective if they were 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 traits such as interest in the targeted domain, self-efficacy, etc. [5].

1.1 Research Hypothesis and Impact

We argue that 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 students' 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)—that is, a little bit beyond the student's current level of understanding about a concept, ability to perform a skill, etc. [26]. Following the practice of human tutors, we propose a computational approach to model the ZPD of students who carry out learning activities using a dialogue-based intelligent tutoring system. In particular, we employ a student model to assess students' changing knowledge as they engage in a dialogue with the system. Based on the model's predictions, we define the concept of the "Grey Area", a probabilistic region in which the model's predictive accuracy is low. We argue that this region can be used to indicate whether a student is in (or not in) the ZPD. Poor predictive accuracy may have two sources: (a) lack of enough evidence (data points) upon which the model can make a good prediction. In that case (insufficient data), the grey area would not be a good approach for modeling the ZPD; (b) the student's performance itself is too variable to make reliable predictions. This mirrors what happens with human tutors: 1) human tutors are not sure what the student knows at the start of a tutoring session (or of a new topic) and 2) human tutors may not be able to assess a student's knowledge if the student is not giving consistent answers for a topic.

We do not claim that the grey area depicts or is directly related to the ZPD. Rather, our research hypothesis is that we can use the outcome of the student model (i.e. the fitted probabilities that predict students' performance) to model students' ZPD. The core rationale is that if the student model cannot predict with acceptable accuracy the performance of a student (i.e. whether a student will answer a question from the tutor 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. Furthermore, we envision that this approach will impact how ITS developers design learning activities and materials, author tutorial dialogues and provide scaffolding and feedback. Even though here we focus on dialogue-based tutoring systems, we expect that our approach can be generalized and expanded to other kinds of ITSs.

In Section 2 we present work related to identifying the ZPD, such as dynamic assessment and provide an overview of student modeling. In Section 3 we discuss our approach and present the study methods. In Section 4 we present the results and findings. We discuss the evaluation, contribution and impact of our approach in Section 5 and conclude by discussing 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 most well-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*" [26]. This definition of the ZPD points out the importance of appropriate assistance in relation to the learning process and development 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*" [22].

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 [4].

Approaches to identifying or to 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 [4]. Various methods that derive from or build on this notion have been developed for the dynamic assessment (DA) of the learning potential of students (or learners in general) [24]. Usually these approaches employ tests that measure the difference between unmediated and mediated performance [20] 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) [12]. However, Dynamic Assessment focuses on assessing the learning or development potential of the learner rather than the actual level of development.

2.2 Intelligent Tutoring Systems and Student Modeling

Intelligent Tutoring Systems (ITSs) commonly use a student model to track the performance of students and subsequently 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 on a concept until she has mastered it almost to certainty. Only then will the student be guided to move forward to other concepts [1, 8, 9]. Mastery learning is in line with the notion of learning curves that is, how many opportunities a student needs in order to master a skill or knowledge component. One could argue that mastery learning is consistent with ZPD theory, in the sense that the student is considered to have mastered a skill when the student is able to successfully carry out, without help, a task that requires this particular skill. However, the ZPD does not directly address mastery but rather potential "development" under appropriate assistance; as aforementioned, 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 [4, 14].

Human tutors do not carry out detailed diagnoses of student knowledge and their assessments of students' knowledge deficits are often inaccurate [7, 21]. However, they nonetheless construct and dynamically update a normative mental representation of students' grasp of the domain content under discussion, as reflected in tutors' adaptive responses to students' need for scaffolding or remediation [18]. It follows logically that a tutorial dialogue system similarly needs 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 subdialogues 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 that 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.

3. METHODS

3.1 Rimac: A Dialogue Tutor for Physics

In this study, we used data collected during three previous studies with the Rimac system to train a student model and test our research hypothesis. Rimac is a web-based natural-language tutoring system that engages students in conceptual discussions after they solve quantitative physics problems [19] and has been used successfully to teach physics concepts to high-school student [16, 17].

The three studies were conducted within high school physics classes at schools in the Pittsburgh, PA area and they followed a similar protocol. First, students took a pretest and were introduced to Rimac. Then they interacted with Rimac to discuss the physics conceptual knowledge associated with quantitative problems on dynamics. Finally, students took a post-test to measure learning differences after interacting with the system. 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 [11]. 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 provides an incorrect answer, 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 on how Rimac is implemented, see [15]). 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.

3.2 The Student Model

For this study, we used an Additive Factor Model (AFM) to model students' knowledge. This student modeling method was introduced into ITS research by Cen et. al. [2, 3]. In this paper we implemented the AFM model following the approach of Chi et al.

[6] since it has been used successfully before to model students who work on physics problems using a dialogue-based tutor. The model predicts the probability of a student completing a step correctly as a linear function of student parameters, knowledge components or skill parameters, and learning. AFM takes into account the frequency of prior practice and exposure to skills but not the correctness of responses.

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:

"*Let`s consider three conceptual questions that are related to this problem and will help you understand the arrow`s motion. 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 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% rule: 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 (i.e., 20% of the total number of student responses; the rest were used for training the model). We chose this training approach because we wanted to study how the same students represented in the training data would perform on future and sometimes unseen steps, and how their knowledge level adapts with practice.

3.3 The Grey Area and the Study Setup

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 high accuracy whether a student will answer a tutor's question correctly (or not), then it might be the case that the student is in the ZPD.

To predict correctness of students' responses in the test set, we used the AFM student model (described in 3.2). Then, we classified the outcome (as correct or incorrect) based on the fitted probabilities provided by the model. In this study, the student model provides 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 it was validated by the ROC curve for the binary classifier (Figure 1). For example, if the fitted probability for a step in the dialogue provided by the model 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 provided by the model is 0.2 (below 0.5) then we expect that the student will not be able to answer the corresponding dialogue step correctly; hence it is classified as incorrect.

We show that the prediction 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 in the pre and post-tests. Similarly, the model will provide low probabilities of correctness (i.e., a low probability to answer a question correctly) for students who performed poorly in the pre and post-tests. By showing that there is a correlation between students' performance in the pre and post-tests and prediction probabilities, we argue that prediction probabilities are appropriate indicators of the ZPD. In this study the pre and post-tests were developed to assess conceptual knowledge associated with the questions and problems that students were assigned to work on, within Rimac.

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. Based on our hypothesis, this window where the prediction error is high can be used to approximately model the student's zone of proximal development. Henceforth we will refer to this window of uncertainty as the "Grey Area". The concept of the Grey Area is depicted in Figure 2 and is on students' performance at the step-level.

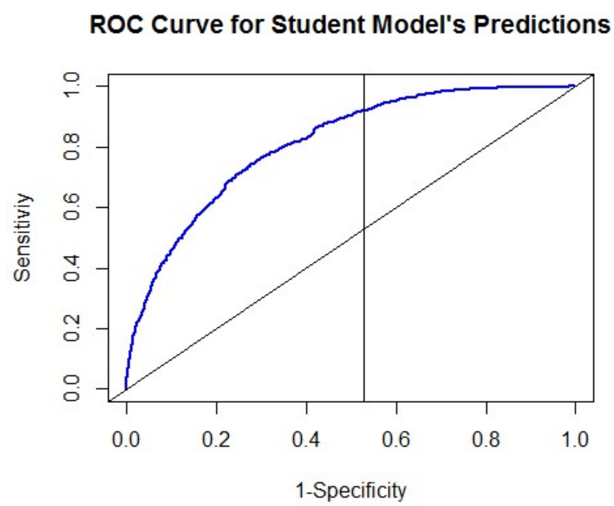

Figure 1 ROC Curve for validating the classification threshold

The space "Above the Grey Area" denotes the area 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 carry out a task 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. We do not propose a specific size but rather try out 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 mainly 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 that it is very important to elicit an answer even if the student is predicted not to be knowledgeable about the topic or a topic might be of minor importance and therefore even a low probability of correctness would be considered sufficient to classify the student as knowledgeable.

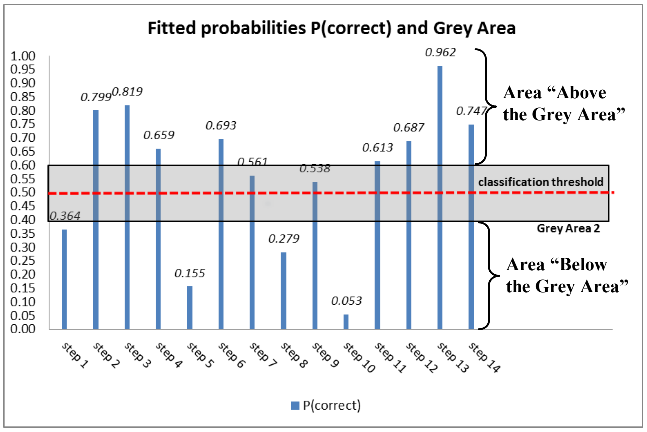

Figure 2. The Grey Area concept with respect to the fitted probabilities (i.e., the probability a student answers correctly on a particular step) as predicted by the student model for a random student and for the various steps of a learning activity. Here we depict the Grey Area ranging from 0.4 to 0.6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

4. ANALYSIS AND RESULTS

4.1 Model Behavior and Student Performance

The research hypothesis is based on the rationale that the prediction 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. The reason for this is that the fitted probability represents the probability that a student will correctly answer a question in the dialogue. Since high performers have a higher probability of answering questions correctly, the average of their fitted probabilities will be higher than those of low performing students.

We performed a correlation analysis to explore this hypothesis. In particular, 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 indicates that with a student who gets a pre-test score revealing she knows a particular KC, the model will predict that the student is able to answer a question that deals with this KC. Similarly, a student who has been classified as able to answer a question that deals with a particular KC will get a post-test score that shows she is knowledgeable about this KC. Conversely, with a student who gets a pre-test score revealing she has no or little knowledge related to a particular KC the model will predict that the student is not able to answer a question that deals with this KC. Subsequently, a student who is classified as unable to answer a question on a particular KC will be expected to score a low 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.

4.2 Model Accuracy for cases inside the Grey Area

As aforementioned, the grey area is defined as the area where the model cannot predict with accuracy 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 size:

- Area 1: contains the fitted probabilities between 0.45 and 0.55
- Area 2: contains the fitted probabilities between 0.4 and 0.6
- Area 3: contains the fitted probabilities between 0.35 and 0.65
- Area 4: contains the fitted probabilities between 0.3 and 0.7
- Area 5: contains the fitted probabilities between 0.25 and 0.75

We chose these particular areas so as 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 or not. Accuracy is defined as the total number of times (1) the student answered correctly and the model also predicted the student would answer correctly and (2) the student answered incorrectly and the model also predicted the student would answer incorrectly divided by the total number of predictions.

The results of this analysis are presented in Table 1 and Table 2. Table 1 presents the analysis of the cases that are contained only in the focal grey area under study (non-cumulative results) and excludes 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. Table 2 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 (Table 1) (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 0.6. As we move away from the classification threshold (0.5), the number of the 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 Figure 3. That finding was expected since the confidence of the model increases. 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 Figure 4 we depict the results for the cumulative analysis. As expected, more cases are predicted correctly as the area size 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". However this work does not aim to define the appropriate size for the Grey Area but rather to study how the model's behavior may change 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%.

Table 1. Predictions' accuracy within grey areas of different sizes – non-cumulative results. Successive areas do not contain cases that were present in preceding areas (e.g., statistics for Area 2 do not take into account the cases contained in Area 1)

Non-Cumulative	Area 1	Area 2	Area 3	Area 4	Area 5
#Cases	394	420	404	369	304
Cases (%)	12.6	13.4	12.9	11.8	9.7
#Correct	213	241	259	250	221
#Incorrect	181	179	145	119	83
Correct (%)	54.1	57.4	64.1	67.8	72.7
Incorrect (%)	45.9	42.6	35.9	32.3	27.3

Table 2. Predictions' accuracy within grey areas of different sizes – cumulative results. Successive areas contain cases that were also present in preceding areas (e.g., statistics for Area 2 also take into account the cases contained in Area 1)

Cumulative	Area 1	Area 2	Area 3	Area 4	Area 5
#Cases	394	814	1218	1587	1891
Cases (%)	12.6	26.0	38.9	50.7	60.4
#Correct	213	454	713	963	1184
#Incorrect	181	360	505	624	707
Correct (%)	54.1	55.8	58.5	60.7	62.6
Incorrect (%)	45.9	44.2	41.5	39.3	37.4

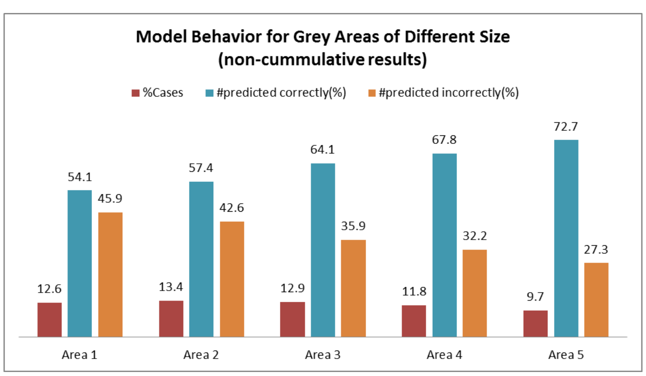

Figure 3. Model behavior (percentage of total number of predicted cases, of cases predicted correctly and cases predicted incorrectly) within the five grey areas of different sizes. The areas are ordered from the narrowest (Area 1) to the widest (Area 5). Each area contains cases that are not contained in preceding areas (non-cumulative analysis).

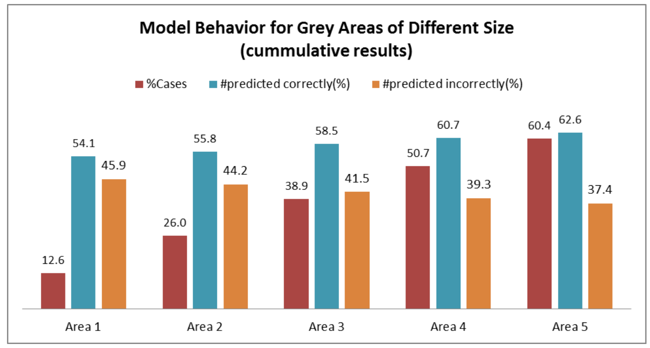

Figure 4. Model behavior (percentage of total number of predicted cases, cases predicted correctly and cases predicted incorrectly) within the five grey areas of different sizes. The areas are ordered from the most narrow (Area 1) to the widest (Area 5). Each area contains cases that are also contained in preceding areas (cumulative analysis).

4.3 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. To that end, it would be interesting to explore the distribution of correct vs. incorrect student answers and whether this distribution would change for grey areas of different sizes. One could argue that based on the way the grey zone was modeled—that is, symmetrical around the cutoff threshold of the binary classified—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.

Table 3. Distribution of correct and incorrect student answers in grey areas of different sizes.

	Area 1	Area 2	Area 3	Area 4	Area 5
#Correct Answers	184	421	633	836	1010
Correct (%)	46.7	51.7	52	52.7	53.4
#Incorrect Answers	210	393	585	751	881
Incorrect (%)	53.3	48.3	48.0	47.3	46.6
Ratio(Cor/Incor)	0.9	1.1	1.1	1.1	1.2
#Cases	394	814	1218	1587	1891

The results are presented in Table 3. The table shows that the percent of correct answers increases as the area widens and that, except for Area 1, the percent of correct answers is larger than the percent of incorrect ones.

In Figure 5 we present the distribution of correct and incorrect answers over grey areas of different sizes and over correct and incorrect model predictions (as shown in the cumulative analysis in paragraph 4.2, Figure 4). For example, for Area 1 the model predicted 54.1% of the cases correctly (e.g., 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 (e.g., the model predicted that a student would answer correctly but the student answered incorrectly, or the model predicted that a student would answer incorrectly but the student answered correctly). 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.

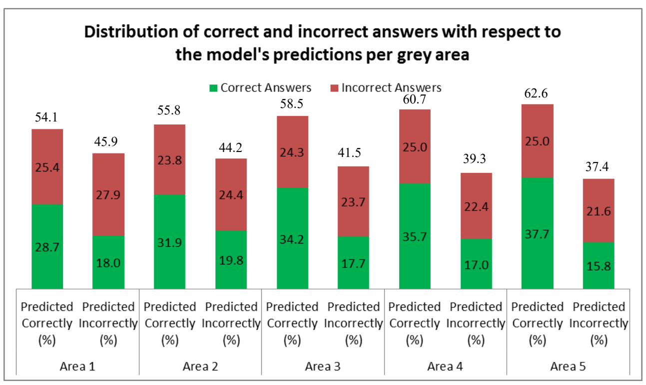

Figure 5. Graphical representation of the distribution of correct and incorrect answers (percentage) with respect to the model's correct and incorrect predictions.

5. DISCUSSION

5.1 Contribution of the approach

We envision that the contribution of the proposed approach, besides its novelty (to the best of our knowledge there is no quantified operationalization of the ZPD) 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. However, this might 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 bored and not challenged – if the task is too easy.

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 systematic way to ensure that a student is (or is not) in the ZPD. One way 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 Grey Area concept). 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:

1. 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;
2. Students who are within the Grey Area will receive meaningful information, scaffolding and hints related to the step in question;
3. 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 that aims to model the Zone of Proximal Development in ITSs. To that end, we introduce the concept of the "Grey Area". It is important to point out that we do not claim that the Grey Area is or can be perceived as 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 is in the ZPD.

To justify our reasoning, 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 prediction 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 indicates that the student model predictions can provide reliable indicators of students' performance.

A limitation of our work is that we have not yet been able to conduct a rigorous evaluation of our approach; however, plans to validate our modeling methods are being developed. Our immediate plan is to carry out extensive studies to explore the proposed approach to modeling the ZPD further, as well as to better understand the strengths and limitations of using a student model to guide students through adaptive lines of reasoning.

7. REFERENCES

[1] Anderson, J.R. et al. 1995. Cognitive tutors: Lessons learned. *The journal of the learning sciences*. 4, 2 (1995), 167–207.

[2] Cen, H. et al. 2008. Comparing two IRT models for conjunctive skills. *International Conference on Intelligent Tutoring Systems* (2008), 796–798.

[3] Cen, H. et al. 2006. Learning factors analysis–a general method for cognitive model evaluation and improvement. *International Conference on Intelligent Tutoring Systems* (2006), 164–175.

[4] Chaiklin, S. 2003. The zone of proximal development in Vygotsky's analysis of learning and instruction. *Vygotsky's educational theory in cultural context*. 1, (2003), 39–64.

[5] Chi, M. et al. 2011. An evaluation of pedagogical tutorial tactics for a natural language tutoring system: A reinforcement learning approach. *International Journal of Artificial Intelligence in Education*. 21, 1–2 (2011), 83–113.

[6] Chi, M. et al. 2011. Instructional factors analysis: A cognitive model for multiple instructional interventions. (2011).

[7] Chi, M.T. et al. 2004. Can tutors monitor students' understanding accurately? *Cognition and instruction*. 22, 3 (2004), 363–387.

[8] Corbett, A.T. et al. 1997. Intelligent tutoring systems. *Handbook of human-computer interaction*. 5, (1997), 849–874.

[9] Corbett, A.T. and Anderson, J.R. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*. 4, 4 (1994), 253–278.

[10] Di Eugenio, B. et al. 2002. The DIAG experiments: Natural language generation for intelligent tutoring systems. *INLG02, The Third International Natural Language Generation Conference* (2002), 120–127.

[11] Evens, M. and Michael, J. 2006. *One-on-one tutoring by humans and computers*. Psychology Press.

[12] Feuerstein, R. et al. 1987. Prerequisites for assessment of learning potential: The LPAD model. (1987).

[13] Graesser, A.C. et al. 2001. Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education*. 12, 3 (2001), 257–279.

[14] Hedegaard, M. 2005. 10 The zone of proximal development as basis for instruction. *An introduction to Vygotsky*. (2005), 227.

[15] Jordan, P. et al. Exploring Contingent Step Decomposition in a Tutorial Dialogue System.

[16] Jordan, P. et al. 2015. Exploring the Effects of Redundancy within a Tutorial Dialogue System: Restating Students' Responses. *16th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (2015).

[17] Jordan, P. et al. 2015. When Is It Helpful to Restate Student Responses Within a Tutorial Dialogue System? *International Conference on Artificial Intelligence in Education* (2015), 658–661.

[18] Katz, S. et al. 2003. Going beyond the problem given: How human tutors use post-solution discussions to support transfer. *International Journal of Artificial Intelligence in Education*. 13, 1 (2003), 79–116.

[19] Katz, S. and Albacete, P.L. 2013. A tutoring system that simulates the highly interactive nature of human tutoring. *Journal of Educational Psychology*. 105, 4 (2013), 1126.

[20] Poehner, M.E. and Lantolf, J.P. 2013. Bringing the ZPD into the equation: Capturing L2 development during Computerized Dynamic Assessment (C-DA). *Language Teaching Research*. 17, 3 (2013), 323–342.

[21] Putnam, R.T. 1987. Structuring and adjusting content for students: A study of live and simulated tutoring of addition. *American educational research journal*. 24, 1 (1987), 13–48.

[22] Reber, A.S. 1995. *The Penguin dictionary of psychology*. Penguin Press.

[23] Rosé, C.P. et al. 2001. Interactive conceptual tutoring in Atlas-Andes. *Proceedings of AI in Education 2001 Conference* (2001), 151–153.

[24] Tzuriel, D. 2000. Dynamic assessment of young children: Educational and intervention perspectives. *Educational Psychology Review*. 12, 4 (2000), 385–435.

[25] VanLehn, K. 2011. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*. 46, 4 (2011), 197–221.

[26] Vygotsky, L. 1978. Interaction between learning and development. *Readings on the development of children*. 23, 3 (1978), 34–41.