Reconciling Adaptivity and Task Orientation in the Student Dashboard of an Intelligent Language Tutoring System

Leona Colling and **Tanja Heck** and **Detmar Meurers**
Department of Computational Linguistics
University of Tübingen
Germany
{leona.colling,tanja.heck,detmar.meurers}@uni-tuebingen.de

Abstract

In intelligent language tutoring systems, student dashboards should display the learning progress and performance and support the navigation through the learning content. Designing an interface that transparently offers information on students' learning in relation to specific learning targets while linking to the overarching functional goal, that motivates and organizes the practice in current foreign language teaching, is challenging. This becomes even more difficult in systems that adaptively expose students to different learning material and individualize system interactions. If such a system is used in an ecologically valid setting of blended learning, this generates additional requirements to incorporate the needs of students and teachers for control and customizability.

We present the conceptual design of a student dashboard for a task-based, user-adaptive intelligent language tutoring system intended for use in real-life English classes in secondary schools. We highlight the key challenges and spell out open questions for future research.

1 Introduction

Language learning is a complex, multidimensional process. It is therefore desirable to provide scaffolding support to learners during practice. Intelligent Tutoring Systems (ITS) can implement means for this purpose in an adaptive way and provide students with insights on their progress and performance (Phobun and Vicheanpanya, 2010).

ITS can accommodate individual differences through macro-adaptive exercise selection and provide micro-adaptive support while working on a selected exercise (Slavuj et al., 2017). Macro-adaptive systems therefore automatically determine the order in which learning content is presented, usually based on a static domain model by matching it to learner characteristics such as proficiency and learning styles (Hafidi and Bensebaa, 2014). Each student receives different learning material which they process at their own pace. The number of exercises a student practices is initially unknown and estimated dynamically after each exercise based on ad-hoc calculations of the student's mastery of the learning object (Rus et al., 2014). Micro-adaptivity, on the other hand, implies that there is no static learning content. Instead, the exercise contents such as hints are dynamically adjusted in order to gradually and individually guide each student towards the correct answers (Lim et al., 2023). Thus, adaptivity improves learning outcomes by adapting to the students' individual needs (Phobun and Vicheanpanya, 2010). Most implementations assign profiles to learners which they generate from training data. Fully adaptive systems then take over all decisions, including structuring and adjusting the learning material based on the learner's profile. This can, however, inhibit them from developing their own learning strategies (Howell et al., 2018). Enabling students to actively engage in the learning decision making process is important to facilitate self-regulation and thus can foster motivation and improve learning outcomes (Lim et al., 2023). Self-regulation can be understood as the students' ability to organize and monitor their own learning behavior and goals by actively managing and shaping their learning environment, such as selecting the next practice target (Schunk and Zimmerman, 2013).

For users to make informed decisions, it is important to show them their personal learning state, according to their interactions with the learning material. Student dashboards generally aim to display information relevant to the student in order to allow them to observe and regulate their learning process. In addition, they provide means to navigate through the learning content (Bull and Kay, 2010). Navigational support is especially relevant and feasible in adaptive systems that incorporate systematically generated, highly variable exercises, such as those following the implementation by Heck and

Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023), pages 288–299
July 13, 2023

Meurers (2022). Both information presentation and navigational structure should be provided in a comprehensible and accessible way. This is particularly important when tailoring a system towards young learners, who are still developing their graphical literacy skills, the ability to understand information presented in graphical form (Roberts and Brugar, 2017).

In order to embed individualized adaptive practice with an ITS into real life, task-oriented language learning classrooms put an important additional focus of the student dashboard on linking practice exercises to their overarching functional goal (Andersen, 2019) and integrating with the curriculum the students follow (Phillips et al., 2020).

In systems used for blended learning in school settings, student dashboards must navigate through the content in a way that aligns with the curriculum, through the systems' default sequence or a sequence defined by the teacher, while maintaining enough flexibility to adjust to students' learning preferences.

In addition, teachers need control over certain aspects of the learning material to satisfy the needs of teacher-guided instruction and successfully combine with the classroom-based teaching (Burstein et al., 2012). Controlling the practiced exercises to a certain extent enables them to refer to the material seen by all of their students in subsequent classroom sessions (Feng et al., 2014). Teachers also want to be able to assign deadlines by which students need to complete practice of certain topics (Hertz, 1992).

Since a curriculum-aligned, structured view of the entire learning content conflicts with the adaptive, dynamic content tailored to the student, it is not straightforward to combine both in a single system. We present an approach to address this challenge by supporting multiple navigational strategies and proposing metrics to display progress and performance overviews which take into account the issues faced by traditional metrics with respect to the demands imposed by adaptivity. Specially tailored towards foreign language learning, our dashboard is co-designed with teachers to keep real life implications in mind and support educational practices when integrated into an Intelligent Language Tutoring System (ILTS) for the use in English classes of secondary schools in Germany.

2 Related work

An increasing number of ITS integrate student dashboards in form of Open Learner Models (OLM) to expose the users to their learning statistics gathered by the system (Bull et al., 2016). This approach has mainly been applied to higher education (Schwendimann et al., 2017), thus not focusing on the particular requirements of systems used in blended learning settings of secondary school teaching. A noticeable exception constitutes the implementation by Rudzewitz et al. (2019) which, however, does not incorporate a task-oriented embedding of the learning content and lacks sufficient simplicity of the visualizations necessary to guide young learners in their self-regulated learning process.

Since most schools nowadays use task-based teaching approaches for language learning (Andersen, 2019), it is necessary to further adapt student dashboards and OLMs to this concept. In order to represent student progress for the various skills practiced in preparation for the functional target task (Ellis, 2016; Mislevy et al., 2002), the dashboard needs to make these task-essential skills explicit to students. Criterion-referenced feedback, which measures performance against predefined criteria, has been successfully explored and evaluated to this purpose (Mirmakhmudova, 2021; Alawar and Abu-Naser, 2017) and later been integrated into an existing ITS by Colling et al. (2022). Their implementation is tailored towards secondary school children by making the visualizations more accessible for the target age group and incorporating task orientation into the dashboard. To this avail, they highlight the functional goal and group exercises and their performance metrics based on curricular units. This contrasts OLMs, which consider the learning domain as a whole (Bull and Kay, 2010). However, their system is not user-adaptive apart from providing scaffolding feedback so that the student dashboard does not consider the requirements introduced by adaptivity.

Integration of macro-adaptive features into a student dashboard depends on the macro-adaptive strategy the system implements. Knowledge Tracing (KT) approaches keep detailed learner models representing the students' progress for various skills within the practiced domain (Liu et al., 2021) and therefore have the benefit of providing progress metrics for the skills which can be made transparent to students in the form of progress bars

(Effenberger, 2018). They do, however, require large amounts of exercises completed by students to train the underlying model (Chen et al., 2018). Since training data for our target group is not readily available, we cannot reliably determine precise progress values. Other approaches use fixed lengths for exercise sequences with incorrect exercises repeated at the end and merely adapt the required complexity of the exercises to select (e.g., Musa and Mohamad, 2017). The progress bar is then only updated when an exercise is solved correctly. While all macro-adaptive systems adaptively determine exercise sequences within a learning object, they pursue varied strategies to determine the order of learning objects. Depending on the degree of self-regulation a system incorporates, it either (a) dictates the entire learning path for the topic to be practiced (Brusilovsky, 1992), (b) requires the learner to choose the next learning object themselves (Twigg, 2003), or (c) provides navigation support without directly enforcing any specific order (Brusilovsky, 2000). As micro-adaptivity changes the exercise content dynamically while students work on it, assigning fixed complexity scores to exercises becomes unreasonable. Macro-adaptive systems therefore typically do not focus on micro-adaptive strategies, apart from providing scaffolded feedback on all exercises.

The body of research on student-facing progress and performance visualizations applicable in adaptive ITS is growing (e.g., Xia et al., 2019; Loboda et al., 2014; Bull and Kay, 2007). Yet, most of these target higher education and thus do not consider the particular needs of teachers and students in schools. Notable exceptions can be found in the domain of mathematics education (e.g., Long and Aleven, 2017). However, to the best of our knowledge, research on student dashboards in adaptive ILTS especially focusing on the demands and needs of ecologically valid K-12 second language learning classrooms is lacking.

With our user-centered design we want to address this gap and offer an approach for a task-oriented student dashboard supporting different navigational strategies in a system simultaneously implementing macro- and micro-adaptivity and used for secondary school English teaching.

3 Dashboard design

Our student dashboard, illustrated in Figure 1, extends a task-oriented dashboard view so that it can be used in an adaptive ILTS supporting teachers in a blended learning context. Where task orientation and adaptivity requirements clash, special considerations are required. The implementation is based on the assumption that most students use the system on a tablet device in landscape mode. This assumption is backed by observations from real life classrooms. The new dashboard features have been co-designed with English teacher practitioners to ensure initial validity. Following the first three stages of the LATUX workflow (Martinez-Maldonado et al., 2015), based on a needs analysis and iterative interviews with teachers using a low-fidelity prototype, we identified requirements on adaptive ILTS used in blended learning with seventh graders and created a high-fidelity prototype with mock learner data for the proposed learner dashboard. The resulting dashboard is described in the following.

Structure The dashboard (see Figure 1) depicts learning content represented in learning units. In accordance with task-based language teaching (Van den Branden, 2016), each unit contains multiple learning targets for grammar or vocabulary practice (e.g., *Simple Past*) which the students need to acquire in order to successfully complete the final communicative target task and its functional goal. In our system, teachers can self-assemble these learning targets into learning units to align with a curriculum, thus supporting different textbooks. Additionally, teachers can define and describe the communicative goal and target task of each learning unit (e.g., *Storytelling: Write a Story! Start with events in the past, describe the present, and then look into the future.*), which will be presented in the dashboard header. Making this link transparent for the students in this way strengthens the connection to the functional target and purpose of practice. Within each learning target of a learning unit, a range of pedagogically motivated realizations of the learning target are listed. *Yes/no questions*, for instance, constitute a realization of the learning target *simple past*. The realizations represent the task-essential language. The system inherent domain model maintains a static, pedagogically motivated order of the learning targets, as well as of the realizations within each target, that have been manually determined by an expert teacher. This structure of the content into coarse and fine-grained content containers makes intermediate acquisition goals visible at different levels

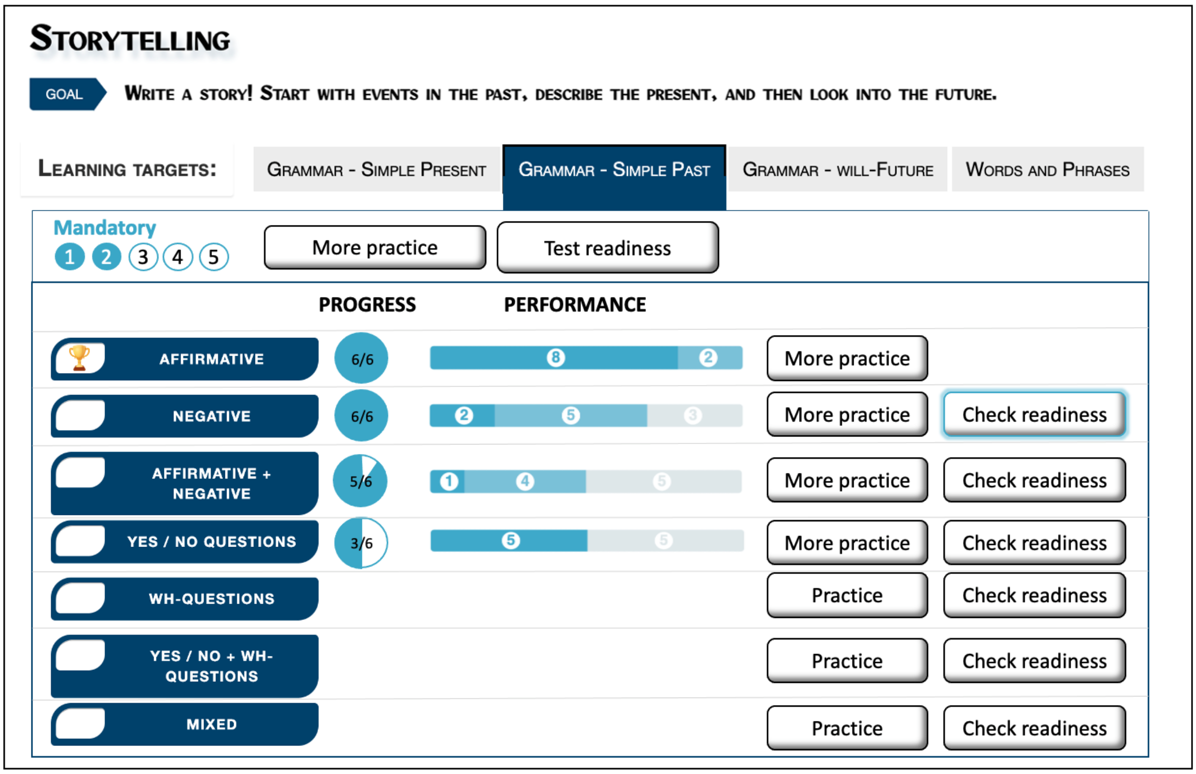

Figure 1: Student dashboard for a user-adaptive, task-oriented ILTS

towards the overarching functional target, which is present at all times, and thus incorporates task-orientation into a student dashboard.

The domain model in our system is based on the curriculum of seventh grade German academic track schools and currently contains 14 grammatical targets, mainly focusing on tenses (i.e., Simple Present, Simple Past, Present Perfect, Past Progressive, Will-Future, Going-to-Future), and their pairwise comparisons, as well as conditional clauses, relative clauses, and comparative forms.

Navigation Traditional systems expect high self-regulation from students by requiring them to themselves navigate through the practice material (Sun et al., 2023). Especially for weaker students, these decisions surpass their abilities so that they do better with adaptive systems (Vandewaetere and Clarebout, 2010). In order to support heterogeneous classrooms with both strong and weak students with different navigational preferences, we integrate a hybrid approach, providing options for less and more learner control over the learning content to practice. In the adaptive practice phase, neither students nor teachers choose distinct exercises, this is done by the adaptive exercise sequencing algorithm. The scope of adaptivity varies depending on the entry point a student chooses. Highly self-regulated students may navigate more autonomously and by themselves select a realization for which the algorithm adaptively sequences exercises. Less self-regulated students, on the other hand, may let the system globally choose both the learning target realization and the exercise.

Although the order displayed via the interface reflects the static order of the pedagogically motivated domain model, the fully adaptive sequence may skip practice of certain realizations for strong students if that realization is also practiced together with other realizations in the same learning target. In the example given in Figure 1, this could for instance be the case for *negative* statements which are also practiced in *affirmative + negative*. Whether a student belongs to the group of strong students for whom realizations are skipped, is based on the student's language proficiency level, which is determined by C-tests periodically administered via the system.

Progress and performance metrics A student dashboard serves not only to navigate to the next exercise but also to visualize the student's progress and performance. Given the lack of sufficient training data for our target domain, we cannot use

KT model-based **progress** representations as commonly used in adaptive systems. In traditional systems, progress can be as simple as displaying the ratio of completed exercises out of all exercises (Duan et al., 2010). This is not suitable for macro-adaptive systems, where the number of completed exercises can easily be determined, yet the number of all exercises is unknown before the student has achieved mastery. Bull and McEvoy (2003) suggest an alternative approach which displays the numbers of successfully and unsuccessfully acquired concepts.

We build our progress metric on this idea, but only display successfully acquired linguistic properties for each realization in a pie chart in order to further increase simplicity. Linguistic properties such as *regular verb forms* are defined at a fine-grained linguistic level. Exercises are linguistically analyzed with the annotation pipeline introduced in Rudzewitz et al. (2018) using the Unstructured Information Management Architecture (UIMA, (Ferrucci and Lally, 2004)) and standard natural language processing (NLP) tools, i.e., segmentation, part-of-speech tagging and dependency parsing with ClearNLP (Choi and Palmer, 2012), lemmatization with Morpha (Minnen et al., 2001) and morphological analysis with the Sfst tool (Schmid, 2005). Based on these basic linguistic analyses of the exercise content, including the target answer and any linguistic co-text such as prompts but excluding exercise instructions, the exercise annotations are extended with more specific linguistic constructions (e.g., *regular verb forms with infinitive ending in -y*) they cover. This second step uses a rule-based approach with UIMA Ruta (Kluegl et al., 2016) as described by Quixal et al. (2021). The domain model hierarchically associates linguistic constructions with properties, and properties with realizations. Thus, it indirectly links annotated exercises to realizations for which they act as options for adaptive practice. Acquisition of these properties represents discrete steps towards progress completion for a realization. Progress completion is calculated based on interactions with the exercises and pre-defined accuracy thresholds per property. Students' attempts on exercise items are analyzed with respect to correctness, therefore a student's answer is compared to the underlying exercise's target answer. Given that the exercise carries annotations of linguistic property, the interactions with items in the exercise result in either positive or negative evidence for property acquisition.

Micro-adaptive adjustments while a student works on an exercise, for instance reducing the number of distractors, are not explicitly shown in the dashboard. They are implicitly incorporated in the progress metric as the adaptive algorithm takes the support a student needs into account by weighting the student's attempts respectively.

In existing ITS, the **performance** achieved for a realization is often indicated based on a single exercise, be it the most recent (e.g., Harindranathan and Folkestad, 2019; Britain, 2020) or the best one per realization-inherent difficulty level (e.g., Colling et al., 2022). In our adaptive system, neither of the two makes much sense. Displaying the performance on a single exercise only makes sense if all exercises target similar properties. Since in our implementation, each realization practices various linguistic properties which are distributed over multiple exercises, a single exercise cannot be representative of a student's current performance. Other systems use average performance over all exercises (Keleş et al., 2009). In this approach, performance visualizations of 100% can only be achieved if all answers are correct. However, students might initially provide incorrect answers based on learning gaps or misconceptions which they can overcome in the practice phase. Displaying average performance of all exercises carries the risk of demotivating or even frustrating students as they cannot receive a perfect performance once given a single incorrect answer. In our system, we want to encourage students to also attempt exercises that they cannot master at first try, to benefit from the scaffolding feedback. Pushing students to only work on exercises where they are certain to get everything correct, in order to have a perfectly polished dashboard with 100% in all performance metrics, would be counter-productive for the purpose of learning and practicing in the zone of proximal development, which describes the space of what a learner can acquire when supported (Vygotsky, 1978). Moreover, average performance is not comparable across students and learning target realizations as it does not account for the amount of practice. A metric based on three exercises would put more weight on incorrect solutions than a metric based on 50 exercises. Average values, given in percentages, in general make it less transparent and less intelligible for low literate students to connect the exercise submission to the performance metrics.

To account for compatibility, transparency and taking learners improvement over time into account, we determine performance by including the most recent ten items instead of focusing on a single submission or an average for an aggregated visualization. As exercises in our system consist of five items, this represents the performance on the last two exercises. Performance is displayed as criterion-referenced performance in a stacked bar chart, giving discrete numbers of items solved correctly at first try, correctly after feedback, and incorrect or not attempted items. This performance display proposed by Colling et al. (2022) shows independent, exam-like as well as scaffolded success and has been evaluated in terms of comprehensibility for seventh graders.

Mastery criterion In order to complete the entire learning target, students need to master all its realizations. Mastery is assessed through specific exercises, which we call *diagnostic exercises*. These are manually created by teachers and didacticians and tailored to align well with the practice exercises and the German seventh grade curriculum. Following Colling et al. (2022)'s approach of parallel exercises, there are multiple comparable instances of diagnostic exercises for each realization. This allows students to re-attempt the readiness check after failing a diagnostic exercise. The current diagnostic exercise for a realization is accessible via the `Check readiness` button and assesses the abilities needed to support the functional goal and thus the student's readiness for the target task regarding the particular realization. It takes into account that no support is provided in the communicative task and therefore evaluates only the student's unassisted attempts without providing scaffolding feedback. When a student achieves mastery for a realization by successfully completing its diagnostic exercise, that realization is assigned a trophy symbol. In traditional ITS, readiness to attempt the diagnostic exercise would correspond to having completed all practice exercises of the realization. As there is no predefined sequence of exercises in a macro-adaptive system, in our approach, the adaptive algorithm evaluates, while the student progresses through the adaptive sequence, if the student has practiced all linguistic properties that underlie the realization and if the student's accuracy is at the required proficiency level. Only then can the system reliably predict that the student will give a correct solution in the diagnostic exercise. Predicted readiness is made salient by a shiny border around the `Check readiness` button in addition to the full progress pie chart. If a student chooses to work on a diagnostic exercise before the system deems them ready, the system advises them to first practice some more, yet without forcing them to do so. Students are thus guided and scaffolded in the understanding of the provided analytics, in form of progress and performance metrics. This enables students to make sense of their statistical data (van Leeuwen et al., 2022) and as a result identify the next steps towards their learning goal.

Permanency of mastery Since traditional ILTS have static exercise sequences, mastery is a permanent attribute. In adaptive systems, however, forgetting needs to be incorporated (Zaidi et al., 2020). In order to consider this in the student dashboard, our trophies "gather dust" once the adaptivity algorithm ascertains that mastery has expired, as demonstrated in Figure 2. This happens if a student hasn't actively – as part of a gap in a gap-filling exercise – or passively – as part of the gaps' co-text – practiced the realization for a set time, which is adjusted based on a student's retention capacities tracked in the learner model. By revising a realization through clicking on the `Check readiness` button – after optionally completing additional practice exercises –, students can prove their maintained proficiency to the system and to themselves. The trophy then regains its shiny appearance.

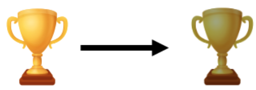

Figure 2: Mastery trophy transition from shiny, i.e., active mastery of respective competence, to dusty, i.e., indicating potential forgetting

Homework assignment Teachers who use an ITS as assistant tool for classroom teaching often desire to assign specific exercises to their students which they can then discuss with the entire class (Singh et al., 2011). This is no problem in traditional systems where all exercises are listed and can directly be accessed by the students. In macro-adaptive systems, however, the algorithm dynamically determines the concrete exercise instance which a student practices. Additionally, in micro-adaptive systems each student receives individually tailored exercises. Consequently, no list of all exercises that is shared amongst all students is available. In order to still facilitate the presented scenario, our system allows teachers to specify mandatory exercises which all students have to complete. By thus giving teachers certain control over the learning content, they can ensure that all students have been exposed to a specific set of hand-selected exercises. The adaptivity algorithm automatically integrates these mandatory exercises into the exercise sequence at the appropriate position according to their associated learning target realizations and the individual student's learning state. However, some students might wish to specifically practice these exercises, either because their slow progress prevents them from reaching the mandatory exercises within the adaptive sequence in a feasible amount of time, or because they do not see the need for additional practice as they might already be proficient in the respective realization. We therefore also explicitly list the mandatory exercises with the option to open them directly. Similarly to the behavior when attempting a diagnostic exercise, if a student chooses to practice a mandatory exercise for which they are not yet proficient enough according to the adaptivity algorithm, the system recommends to first practice more. Students can always decide to ignore these recommendations and proceed to the selected mandatory exercise.

These considerations allow the system to provide common student dashboard features of task-based ITS while also integrating user-adaptivity.

4 Discussion

Since research on simultaneous integration of a task-oriented student dashboard and user-adaptivity in systems applied at secondary school level is very limited, alternative approaches can be considered in some cases and some concepts still lack empirical validation. We therefore discuss potential issues with and alternatives for some of our proposed implementations.

Diagnostic exercises In our approach, students can attempt diagnostic exercises by clicking the `Check readiness` button either next to a learning target realization or globally for the entire learning target comprising all its realizations. On the one hand, this global entry point is in line with the adaptive approach requiring low self-regulation. On the other hand, students using the global button receive all diagnostic exercises in succession so that they do not directly follow the exercises which prepare for them. This decouples the diagnostic exercises for a realization from their scaffolding practice exercises. If the global `Check readiness` button was removed, the question would remain whether all students should proactively attempt the diagnostic exercises themselves – which would potentially result in the same dilemma for students following the adaptive sequence, as they would not be assisted in when to attempt which diagnostic exercise. A solution could be to integrate the diagnostic exercises into the adaptive exercise sequence and saliently flag them for students. Students should then get the choice to attempt the exercise or practice more.

Transparency The subject of mandatory exercises leaves an additional question to be addressed. While displaying them globally for the entire learning target avoids the issue of being inaccessible on-demand, it also removes visual assignment to any realization. Since the aim of explicitly listing the realizations is to also foster meta-linguistic knowledge (Godwin-Jones, 2021), neglecting this aspect for mandatory exercises is questionable. Moreover, this would make it harder for students to autonomously reconstruct progress and performance updates from exercise submissions, thus resulting in higher mental load. The lack of transparency in linking exercises to realizations is also an issue for exercises accessed via the global adaptivity buttons `More practice` and `Check readiness` for the entire learning target. This could potentially be addressed by highlighting the associated realization upon opening the exercise.

Performance visualization A further discussion point concerns the visualization of student performance. While we have presented an approach to display it as criterion-referenced performance on the most recent items, multiple alternative aggregations and visualizations are envisionable. Representing mastery estimates of concepts (Tong et al., 2022), taking the average performance over multiple exercises, adding up the scores for a defined number of items or only displaying those of the most recent exercise are all valid options (Van Labeke et al., 2007; Harindranathan and Folkestad, 2019). Instead of aggregating multiple exercises, the student dashboard could also visualize all completed exercises for a student individually. This would also make it possible to mark mandatory exercises. However, if a student practices a lot, the dashboard could quickly become crowded and therefore poorly accessible on small displays (Bull and Kay, 2016). Reducing the exercises to dot representations might increase manageability. Criterion-referenced performance for each exercise could then be displayed on demand after clicking on a dot. In order to increase the discriminability of dot representations, different colors could indicate certain properties of the exercises such as the exercise type, the exercise dimension, or the proficiency level. Making the exercise type salient could for example address varying complexities as for example gap-filling exercises are considered more complex than multiple-choice exercises (Medawela et al., 2018), or varying foci on language dimensions (Grellet, 1981, p. 5) inherent to different exercise types. It would, however, still fail to consider differences in exercise complexity within each exercise type, for instance based on the number of distractors in multiple-choice exercises (Heck et al., 2022). Using the exercise dimensions of receptive, interactive and productive types (Vetter, 2012) instead would reduce the number of categories and thus increase heterogeneity within each category. Since macro-adaptivity aims to gradually increase exercise complexity, the different categories would for both options inadvertently display scores at different stages of the learner's progress, which might not be transparent to students. The alternative approach to associate exercises with the learner's proficiency level at the time of completing them translates continuous proficiency scores of a KT model into concrete categories. Considering the small number of categories, this is also feasible with KT models of moderate accuracy. However, since a student's progress is not always linear (Shirai, 1990) nor are there clear thresholds between the levels, this approach might not give helpful insights either. A compromise between representing all exercises and using a single global aggregation could alternatively collapse exercises with similar colors into a single dot representation with the number of collapsed exercises indicated inside the dot. This would, however, lose the benefits of the non-collapsed representation of highlighting mandatory exercises and providing anchors for criterion-referenced performances per exercise.

Progress visualization Although we choose to base our progress measure on linguistic properties, this does not necessarily have to be the case. The categories of exercise type, exercise dimension, and exercise complexity suggested for a performance metric can also be considered for progress. However, categorical progress units, which increase in discrete steps, incorporate ranges of continuous values so that progress does not necessarily increase after each exercise. While KT in principle facilitates continuous and constantly perceivable progress updates, we have already argued that the model's estimates are not accurate enough with insufficient training data.

Customized learning units Finally, an adaptive system that supports multiple curricula allows teachers to compile their own learning units. Ideally, teachers can also exclude certain linguistic properties which they do not (yet) wish to practice. Since they may later decide to include these properties, students who have already received a trophy might not fulfill the requirements anymore when also considering the newly included properties. Withdrawing the trophy could be discouraging and the underlying reasoning might not be intuitive to students. A possible solution could use the mechanism of gathering dust so that the trophy would still be visible but inactive. Additionally, the progress pie chart for the realization would have to change accordingly. Making these changes transparent and intelligible for students is not trivial, especially considering that young learners' graphical literacy skills may still be developing (Roberts and Brugar, 2017). It becomes even more of a challenge if multiple learning units practice the same learning target, thus sharing the same pool of exercise candidates. This is especially relevant for teachers who want to incorporate a revision learning target, e.g., having one learning unit where students first learn *simple past*, maybe not including all linguistic properties, but also including *simple past* as a revision when introducing *conditionals type 2* in another learning unit. The question then arises whether performance should be calculated separately within each learning unit – which would hinder the adaptivity algorithm as it would not be able to globally track the students' learning progress – or synchronize progress across the units. Synchronizing progress for realizations where different linguistic properties have been excluded is, however, unfeasible. On the other hand, disallowing teachers to exclude different properties for different units might also not result in the desired functionalities, especially if teachers intend to practice complementary properties of a realization in different learning units. From a student's perspective, working on a learning target realization in one learning unit but receiving a performance update in another unit as well might lead to misunderstandings, demotivation, or distrust in the technology due to the poor user experience (Franconeri et al., 2021). In the worst case, it could even result in negative learning outcomes. From a teacher's perspective, interpreting and assessing duplicates of identical performance history items in multiple units might be challenging and tedious. Especially the display of mandatory exercises in synchronized learning targets constitutes an open issue.

5 Conclusion

We presented the design of a student dashboard for an ITS which integrates curriculum-driven, task-based language teaching and user-adaptivity and has been designed in a co-participatory approach with teachers. We outlined an implementation based on practices and insights from these two instructional approaches that takes into account the opportunities, but also the requirements and restrictions of both. Taking this design as starting point, we critically discussed potential limitations and alternative approaches. Such conceptual and theoretical discussions will guide future work in terms of implementation and evaluation of the dashboard in authentic settings. In a next step, to pilot the design and decide on some open alternatives before fully implementing the dashboard into the system, we plan to evaluate the high-fidelity prototype in a user study with teachers to ascertain efficacy of the design. In this user study we will obtain first quantitative measures on usability and intelligibility. Based on these findings, the refined student dashboard will be implemented using recent front-end development libraries like REACT[1] and build into the modular architecture of the ILTS FeedBook (Parrisius et al., 2022), connecting the dashboard with FeedBook's existing micro-service landscape, including the adaptivity micro-service, the one for NLP processing and the learner model micro-service. The fully implemented dashboard integrated into the ILTS will then be evaluated in a large-scale field study with student participants using the system in a blended learning setting over an extended period of a school year. The data collected in that study will allow us to identify different learning paths and map them to student characteristics such as high and low self-regulation and navigational preferences such as globally adaptive, realization adaptive or completely self-guided sequencing. Furthermore, the study will yield valuable insights into the practicability and acceptability of the design in real-world usage.

[1] https://react.dev/

Acknowledgements

We want to acknowledge the teachers in the AI2Teach project, especially Florian Nuxoll, for their input and feedback.

References

Mariam W. Alawar and Samy S. Abu-Naser. 2017. CSS-Tutor: An intelligent tutoring system for CSS and HTML. *International Journal of Academic Research and Development*, 2(1):94–99.

Katja Andersen. 2019. Assessing task-orientation potential in primary science textbooks: Toward a new approach. *Journal of Research in Science Teaching*, 57:481–509.

Gabriel Wade Britain. 2020. *Design Analytics Dashboards to Support Students and Instructors*. Ph.D. thesis, Texas A&M University.

Peter Brusilovsky. 2000. Adaptive Hypermedia: From Intelligent Tutoring Systems to Web-Based Education. In *Intelligent Tutoring Systems*, pages 1–7, Berlin, Heidelberg. Springer Berlin Heidelberg.

Peter L. Brusilovsky. 1992. A framework for intelligent knowledge sequencing and task sequencing. In *Intelligent Tutoring Systems*, pages 499–506, Berlin, Heidelberg. Springer Berlin Heidelberg.

Susan Bull, Blandine Ginon, Clelia Boscolo, and Matthew Johnson. 2016. Introduction of Learning Visualisations and Metacognitive Support in a Persuadable Open Learner Model. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, LAK '16, pages 30–39, New York, NY, USA. Association for Computing Machinery.

Susan Bull and Judy Kay. 2007. Student models that invite the learner in: The smili:() open learner modelling framework. *International Journal of Artificial Intelligence in Education*, 17(2):89–120.

Susan Bull and Judy Kay. 2010. Open Learner Models. *International Journal of Artificial Intelligence in Education*, 308:301–322.

Susan Bull and Judy Kay. 2016. SMILI: a Framework for Interfaces to Learning Data in Open Learner Models, Learning Analytics and Related Fields. *International Journal of Artificial Intelligence in Education*, 26(1):293–331.

Susan Bull and Adam Thomas McEvoy. 2003. An Intelligent Learning Environment with an Open Learner Model for the Desktop PC and Pocket PC. In U. Hoppe, F. Verdejo, and J. Kay, editors, *Artificial Intelligence in Education*, pages 389–391. IOS Press, Amsterdam.

Jill Burstein, Jane Shore, John Sabatini, Brad Moulder, Steven Holtzman, and Ted Pedersen. 2012. The language musesm system: Linguistically focused instructional authoring. *ETS Research Report Series*, 2012(2):i–36.

Penghe Chen, Yu Lu, Vincent W. Zheng, and Yang Pian. 2018. Prerequisite-Driven Deep Knowledge Tracing. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 39–48, Los Alamitos, CA, USA. IEEE Computer Society.

Jinho D. Choi and Martha Palmer. 2012. Fast and robust part-of-speech tagging using dynamic model selection. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 363–367, Jeju Island, Korea. Association for Computational Linguistics.

Leona Colling, Ines Pieronczyk, Stephen Bodnar, Heiko Holz, Cora Parrisius, Katharina Wendebourg, Carolin Blume, Florian Nuxoll, Diana Pili-Moss, and Detmar Meurers. 2022. Practice with a purpose. development of a task-oriented learner dashboard. EUROCALL Conference 2022, virtual.

Dandi Duan, Antonija Mitrovic, and Neville Churcher. 2010. Evaluating the Effectiveness of Multiple Open Student Models in EER-Tutor. In *Proceedings of the 18th International Conference on Computers in Education*, pages 86–88, Putrajaya, Malaysia. University of Canterbury. Computer Science and Software Engineering.

Tomáš Effenberger. 2018. *Adaptive system for learning programming*. Ph.D. thesis, Master's thesis, Masaryk University.

Rod Ellis. 2016. Focus on form: A critical review. *Language Teaching Research*, 20(3):405–428.

Mingyu Feng, Jeremy Roschelle, Neil Heffernan, Janet Fairman, and Robert Murphy. 2014. Implementation of an Intelligent Tutoring System for Online Homework Support in an Efficacy Trial. In *"Intelligent Tutoring Systems"*, pages 561–566, Cham. Springer International Publishing.

David Ferrucci and Adam Lally. 2004. Uima: an architectural approach to unstructured information processing in the corporate research environment. *Natural Language Engineering*, 10(3-4):327–348.

Steven L. Franconeri, Lace M. Padilla, Priti Shah, Jeffrey M. Zacks, and Jessica Hullman. 2021. The science of visual data communication: What works. *Psychological Science in the public interest*, 22(3):110–161.

Robert Godwin-Jones. 2021. Big data and language learning: Opportunities and challenges. *Language Learning & Technology*, 25(1):4–19.

Francoise Grellet. 1981. *Developing Reading Skills: A Practical Guide to Reading Comprehension Exercises*. Cambridge Language Teaching Library. Cambridge University Press, New York.

Mohamed Hafidi and Tahar Bensebaa. 2014. Developing Adaptive and Intelligent Tutoring Systems (AITS): A General Framework and Its Implementations. *International Journal of Information and Communication Technology Education*, 10(4):70–85.

Priya Harindranathan and James Folkestad. 2019. Learning Analytics to Inform the Learning Design: Supporting Instructors' Inquiry into Student Learning in Unsupervised Technology-Enhanced Platforms. *Online Learning Journal*, 23(3):34–55.

Tanja Heck and Detmar Meurers. 2022. Parametrizable exercise generation from authentic texts: Effectively targeting the language means on the curriculum. In *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*, pages 154–166, Seattle, Washington. Association for Computational Linguistics.

Tanja Heck, Detmar Meurers, and Florian Nuxoll. 2022. Automatic exercise generation to support macro-adaptivity in intelligent language tutoring systems. In *Intelligent CALL, granular systems and learner data: short papers from EUROCALL 2022*, pages 162–167. Research-publishing.net.

Alain Hertz. 1992. Finding a feasible course schedule using tabu search. *Discrete Applied Mathematics*, 35(3):255–270.

Joel A Howell, Lynne D Roberts, Kristen Seaman, and David C Gibson. 2018. Are we on our way to becoming a "helicopter university"? academics' views on learning analytics. *Technology, Knowledge and Learning*, 23:1–20.

Aytürk Keleş, Rahim Ocak, Ali Keleş, and Aslan Gülcü. 2009. Zosmat: Web-based intelligent tutoring system for teaching–learning process. *Expert Systems with Applications*, 36(2):1229–1239.

Peter Kluegl, Martin Toepfer, Philip-Daniel Beck, Georg Fette, and Frank Puppe. 2016. Uima ruta: Rapid development of rule-based information extraction applications. *Natural Language Engineering*, 22(1):1–40.

Lyn Lim, Maria Bannert, Joep van der Graaf, Shaveen Singh, Yizhou Fan, Surya Surendrannair, Mladen

Rakovic, Inge Molenaar, Johanna Moore, and Dragan Gašević. 2023. Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior*, 139:107547.

Qi Liu, Shuanghong Shen, Zhenya Huang, Enhong Chen, and Yonghe Zheng. 2021. A Survey of Knowledge Tracing.

Tomasz Dominik Loboda, Julio Guerra, Roya Hosseini, and Peter Brusilovsky. 2014. Mastery grids: An open-source social educational progress visualization. In *Proceedings of the 2014 conference on Innovation & technology in computer science education*, pages 357–357.

Yanjin Long and Vincent Aleven. 2017. Enhancing learning outcomes through self-regulated learning support with an Open Learner Model. *User Modeling and User-adapted Interaction*, 27(1):55–88.

Roberto Martinez-Maldonado, Abelardo Pardo, Negin Mirriahi, Kalina Yacef, Judy Kay, and Andrew Clayphan. 2015. Latux: An iterative workflow for designing, validating, and deploying learning analytics visualizations. *Journal of Learning Analytics*, 2(3):9–39.

R.M. Sumudu Himesha B Medawela, Dugganna Ralalage Dilini Lalanthi Ratnayake, Wijeyapala Abesinghe Mudiyanselage Udari Lakshika Abeyasinghe, Ruwan Duminda Jayasinghe, and Kosala Nirmalani Marambe. 2018. Effectiveness of "fill in the blanks" over multiple choice questions in assessing final year dental undergraduates. *Educación Médica*, 19:72–76.

Guido Minnen, John Carroll, and Darren Pearce. 2001. Applied morphological processing of english. *Natural Language Engineering*, 7(3):207–223.

Iroda Mirmakhmudova. 2021. Comparing criterion and norm referenced assessments of langauge skills in the second language. *Asian Journal of Social Sciences and Humanities*, 11:463.

Robert J. Mislevy, Linda S. Steinberg, and Russell G. Almond. 2002. Design and analysis in task-based language assessment. *Language Testing*, 19(4):477–496.

Nushi Musa and Hosein Eqbali Mohamad. 2017. Duolingo: A mobile application to assist second language learning. *Teaching English with Technology*, 17(1):89–98.

Cora Parrisius, Ines Pieronczyk, Carolyn Blume, Katharina Wendebourg, Diana Pili-Moss, Mirjam Assmann, Sabine Beilharz, Stephen Bodnar, Leona Colling, Heiko Holz, et al. 2022. Using an intelligent tutoring system within a task-based learning approach in english as a foreign language classes to foster motivation and learning outcome (interact4school): Preregistration of the study design. PsychArchives.

Andrea Phillips, John Pane, Rebecca Reumann-Moore, and Oluwatosin Shenbanjo. 2020. Implementing an adaptive intelligent tutoring system as an instructional supplement. *Educational Technology Research and Development*, 68(3):1409–1437.

Pipatsarun Phobun and Jiracha Vicheanpanya. 2010. Adaptive intelligent tutoring systems for e-learning systems. *Procedia Social and Behavioral Sciences*, 2(2):4064–4069.

Martí Quixal, Björn Rudzewitz, Elizabeth Bear, and Detmar Meurers. 2021. Automatic annotation of curricular language targets to enrich activity models and support both pedagogy and adaptive systems. In *Proceedings of the 10th Workshop on NLP for Computer Assisted Language Learning*, pages 15–27.

Kathryn L. Roberts and Kristy A. Brugar. 2017. The view from here: Emergence of graphical literacy. *Reading Psychology*, 38(8):733–777.

Björn Rudzewitz, Ramon Ziai, Kordula De Kuthy, Verena Möller, Florian Nuxoll, and Detmar Meurers. 2018. Generating feedback for english foreign language exercises. In *Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications*, pages 127–136.

Björn Rudzewitz, Ramon Ziai, Florian Nuxoll, Kordula De Kuthy, and Detmar Meurers. 2019. Enhancing a Web-based Language Tutoring System with Learning Analytics. In *Joint Proceedings of the Workshops of the 12th International Conference on Educational Data Mining co-located with the 12th International Conference on Educational Data Mining, EDM 2019 Workshops, Montréal, Canada, July 2-5, 2019*, volume 2592 of *CEUR Workshop Proceedings*, pages 1–7. CEUR-WS.org.

Vasile Rus, Dan Stefanescu, Nobal Niraula, and Arthur C. Graesser. 2014. DeepTutor: Towards Macro- and Micro-Adaptive Conversational Intelligent Tutoring at Scale. In *Proceedings of the First ACM Conference on Learning @ Scale Conference*, L@S '14, pages 209–210, New York, NY, USA. Association for Computing Machinery.

Helmut Schmid. 2005. A programming language for finite state transducers. In *FSMNLP*, volume 4002, pages 308–309. Citeseer.

Dale H Schunk and Barry J Zimmerman. 2013. Self-regulation and learning. In W. M. Reynolds, G. E. Miller, and I. B. Weiner, editors, *Handbook of psychology: Educational psychology*, pages 45—68. John Wiley & Sons, Inc.

Beat A. Schwendimann, María Jesús Rodríguez-Triana, Andrii Vozniuk, Luis P. Prieto, Mina Shirvani Boroujeni, Adrian Holzer, Denis Gillet, and Pierre Dillenbourg. 2017. Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. *IEEE Transactions on Learning Technologies*, 10(01):30–41.

Yasuhiro Shirai. 1990. U-shaped behavior in L2 acquisition. In H. Burmeister and P. L. Rounds, editors, *Handbook of Japanese Psycholinguistics*, volume 2, pages 217–234. De Gruyter, Eugene, OR.

Ravi Singh, Muhammad Saleem, Prabodha Pradhan, Cristina Heffernan, Neil T. Heffernan, Leena Razzaq, Matthew D. Dailey, Cristine O'Connor, and Courtney Mulcahy. 2011. Feedback during Web-Based Homework: The Role of Hints. In *Artificial Intelligence in Education*, pages 328–336, Berlin, Heidelberg. Springer Berlin Heidelberg.

Vanja Slavuj, Ana Meštrović, and Božidar Kovačić. 2017. Adaptivity in educational systems for language learning: a review. *Computer Assisted Language Learning*, 30(1-2):64–90.

Jerry Chih-Yuan Sun, Hsueh-Er Tsai, and Wai Ki Rebecca Cheng. 2023. Effects of integrating an open learner model with AI-enabled visualization on students' self-regulation strategies usage and behavioral patterns in an online research ethics course. *Computers and Education: Artificial Intelligence*, 4:100120.

Hanshuang Tong, Zhen Wang, Yun Zhou, Shiwei Tong, Wenyuan Han, and Qi Liu. 2022. Introducing Problem Schema with Hierarchical Exercise Graph for Knowledge Tracing. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22, pages 405–415, New York, NY, USA. Association for Computing Machinery.

Carol A. Twigg. 2003. Models for online learning. *Educause review*, 38:28–38.

Kris Van den Branden. 2016. Task-based language teaching. In *The Routledge handbook of English language teaching*, pages 238–251. Routledge, Abingdon, Oxon, UK; New York, NY, USA.

Nicolas Van Labeke, Paul Brna, and Rafael Morales Gamboa. 2007. Opening up the Interpretation Process in an Open Learner Model. *International Journal of Artificial Intelligence in Education*, 17:305–338.

Anouschka van Leeuwen, Stephanie D Teasley, Alyssa Wise, Charles Lang, George Siemens, Dragan Gašević, Agathe Merceron, et al. 2022. Teacher and student facing analytics. In *Handbook of Learning Analytics*, pages 130–140. Society for Learning Analytics Research.

Mieke Vandewaetere and Geraldine Clarebout. 2010. Perceptions and Illusions about Adaptivity and Their Effects on Learning Outcomes. In *Proceedings - 10th IEEE International Conference on Advanced Learning Technologies, ICALT 2010*, pages 480–484, Los Alamitos, CA, USA. IEEE Computer Society.

Eva Vetter. 2012. Exploiting receptive multilingualism in institutional language learning: The case of Italian in the Austrian secondary school system. *International Journal of Bilingualism*, 16(3):348–365.

Lev Semenovich Vygotsky. 1978. *Mind in society: Development of higher psychological processes*. Harvard university press.

Meng Xia, Mingfei Sun, Huan Wei, Qing Chen, Yong Wang, Lei Shi, Huamin Qu, and Xiaojuan Ma. 2019. Peerlens: Peer-inspired interactive learning path planning in online question pool. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, page 1–12, New York, NY, USA. Association for Computing Machinery.

Ahmed Zaidi, Andrew Caines, Russell Moore, Paula Buttery, and Andrew Rice. 2020. Adaptive Forgetting Curves for Spaced Repetition Language Learning. In *Artificial Intelligence in Education*, pages 358–363, Cham. Springer International Publishing.