Towards Task-Oriented ICALL: A Criterion-Referenced Learner Dashboard Organising Digital Practice

Leona Colling[1,6], Ines Pieronczyk[2,6], Cora Parrisius[2,3,6], Heiko Holz[1,4,6], Stephen Bodnar[1,6,7], Florian Nuxoll[1,6] and Detmar Meurers[1,5,6]

[1]*Department of Linguistics, University of Tübingen, Germany*
[2]*Hector Research Institute of Educational Sciences and Psychology, University of Tübingen, Germany*
[3]*Karlsruhe University of Education, Germany*
[4]*Ludwigsburg University of Education, Germany*
[5]*Leibniz-Institut für Wissensmedien (IWM), Germany*
[6]*LEAD Graduate School & Research Network, Germany*
[7]*Tübingen Center for Digital Education, Germany*

Keywords: Second Language Acquisition, Task-Supported Language Teaching, Intelligent Computer Assisted Language Learning, Learner Dashboard, Criterion-Referenced Feedback.

Abstract: Practice is an essential part of learning. Intelligent Computer-Assisted Language Learning (ICALL) systems can provide practice opportunities and give insights into the learner's learning state and progress. Open learner models have been designed to provide learners with information on the overall learning domain. However, current approaches to foreign language teaching typically motivate practice as preparation for a communicative or functional task. This raises the question of how this motivating functional task context and progress towards mastering the task-essential language can be made explicit in an ICALL system. We present an approach to ICALL practice that is orchestrated in a dashboard that provides information on the learner's competence-oriented learning progress towards the overall task goals. The dashboard allows students to choose what to practice next based on this information, which provides a transparent, motivating link to the purpose of practicing. Organising digital practice based on a task- and competence-oriented curriculum also facilitates the acceptance of ICALL in the formal school setting. The dashboard introduced in this article extends the intelligent tutoring system FeedBook for English in German secondary schools. The article provides the theoretical background for the dashboard's structure, motivates the design process, and describes the resulting implementation.

1 INTRODUCTION

While practice has long been an integral part of foreign language learning, Skill Acquisition Theory (DeKeyser, 2020) has scientifically motivated the role and importance of practice in Second Language Acquisition (SLA) research. Systematic practice enables learners to proceduralise and partially automatise language production (DeKeyser, 2010). In this context, new technologies such as Intelligent Computer-Assisted Language Learning (ICALL) systems offer a great opportunity to enhance the learning experience and examine practice behaviour in authentic school contexts (Meurers et al., 2019; Ruiz et al., 2023). ICALL systems can provide exercises for practice with adaptive feedback, track the learner interaction with the system, and display insights on the language learning progress (Rudzewitz, 2021), parallel to what has been established for other learning domains, such as arithmetic (Molenaar & Knoop-Van Campen, 2016). Based on interaction data, the systems can build an internal representation of the learner's knowledge and misconceptions across the subject domain, the learner model. When it is also made accessible to the user, it is often referred to as an Open Learner Model (OLM, Bull, 2004). Bodily et al. (2018) emphasise the beneficial effects of the use of data gathered and visualized in an ICALL system on the learner's reflection and self-regulation, relating it to the field of data-driven learning analytics as well as a pedagogical point of view.

At the same time, the implementation of ICALL in authentic school contexts is very slow (Schmidt &

Colling, L., Pieronczyk, I., Parrisius, C., Holz, H., Bodnar, S., Nuxoll, F. and Meurers, D.
Towards Task-Oriented ICALL: A Criterion-Referenced Learner Dashboard Organising Digital Practice.
DOI: 10.5220/0012753000003693

In *Proceedings of the 16th International Conference on Computer Supported Education (CSEDU 2024) - Volume 1*, pages 668-679
ISBN: 978-989-758-697-2; ISSN: 2184-5026

Strasser, 2022). One aspect holding back its implementation is that exercises generally need to be embedded into a pedagogical sequence of introduction, systematisation, and practice of the language means toward functional language use (Brauner & Prediger, 2018). Indeed, successful participation in an authentic, often communicative task at the end of a learning unit forms the central goal of task-oriented teaching approaches, such as task-supported language teaching (TSLT; Müller-Hartmann & Schocker von Ditfurth, 2011) and task-based language teaching (TBLT; Ellis, 2003; Robinson, 2011; Willis & Willis, 2007). It has become standard in German school curricula for foreign language learning to centre language instructions around such tasks.

Pre-task activities offering practice opportunities can build up towards the task goal (Vogt & Schmidt, 2021). These systematic activities allow learners to practice task-essential language aspects as part of homework or during individual learning times. They provide a natural way of integrating ICALL practice into current task-oriented foreign language classrooms. Technology-mediated TBLT, where technology is used to aid TBLT, has been qualitatively examined (see Chong & Reinders, 2020) and researchers even identified design patterns for task-based technology-enhanced language learning (Canals & Mor, 2020). However, in contrast to 'strong' TBLT, where focus on forms is incidental, the school reality in Germany is more in line with TSLT, where linguistic forms are taught explicitly prior to carrying out a meaningful, contextualised, and interactive task (Kolb & Raith, 2018; Kos 2023). To the best of our knowledge, we here present the first ICALL approach for large-scale use in K-12 classrooms explicitly designed to facilitate the TSLT pre-task practice phase.

This paper discusses how an ICALL system can be extended with a learner dashboard to motivate and embed the practice activities as part of a task-supported curriculum. First, we will introduce OLMs and TSLT as the foundations for our learner dashboard design. Second, we will describe our concrete realisation of a task-based learner dashboard in an ICALL system for English learners in secondary schools.

2 COMBINING OPEN LEARNER MODELS & TASK-SUPPORTED LANGUAGE TEACHING

Computer-assisted language learning (CALL) has a long history dating back to the 1960s (Davies et al., 2013). While interest in CALL rose in the 1990s (Chen et al., 2020), Intelligent CALL (ICALL) is a more recent research strand employing methods from natural language processing and artificial intelligence (Schulze & Heift, 2013). Complementing these methods to analyse language and provide feedback, ICALL approaches realising Intelligent Language Tutoring Systems (ILTS) also maintain learner models to monitor progress and adapt system interaction (Heift & Schulze, 2003). Sometimes this information is also made openly available to the learner (see e.g., Bull & Pain, 1995). After providing some background, we raise the question of how we can organise such information about the learner and the learning process to facilitate interaction in relation to an overarching didactic approach, such as task-supported foreign language teaching as a prominent paradigm.

2.1 Open Learner Models (OLMs)

Bull and Wasson (2016) characterize OLMs as visualizations that are accessible for the learner and are "based on an underlying inferred model of the learner's current competences or understanding, rather than behaviour or performance data logged" (p. 151). This is rooted in a long tradition of learner models as a representation of the learners' knowledge of a subject domain as a whole (Bull & Kay, 2010). Such learner models are a core component of adaptive digital learning environments (Brusilovsky & Millán, 2007) to let the system act on the basis of the inferred student learning status, competencies, or misconceptions (Bull, 2004). Making use of the same type of information, OLMs provide the learners with information about their learning to help them monitor and regulate it (Bodily et al., 2018; Bull & Kay, 2010).

In the SLA context, OLMs have been implemented in some systems, primarily in the higher education context (e.g., Bull et al., 2016; Bull & Wasson, 2016; Tsourounis & Demmans Epp, 2016; Xu & Bull, 2010). To depict the current inferred learning state to the learner, skill meters, gauges, stars, or radar plots have become established visualisations. Learners independently explore these visualisations of their personal learner model representing their knowledge of the overall curriculum. This is complex and may be the reason why there is substantially less work on OLMs with younger learners – though Y. Long & Aleven (2017) is of relevance here, and in Rudzewitz et al. (2020) we presented an OLM for the secondary school English context.

While OLMs provide a wealth of information on competency states and misconceptions on each of the many different subdimensions of the subject domain, it is often not transparent for the learner how completing an exercise contributes to the system's record of the overall learning state. The complexity of the curriculum and how different types of exercises instantiate the relevant language competencies contrasts with the learners' need for easily accessible information. Hard to understand visualisations will fail to impact learning behaviour (Bull, 2004; Ring et al., 2019). Going beyond the need to present the relevant information transparently, OLMs exclusively reporting the students' learning status for the subject domain are missing an organisational and actionable link to the learning goal and context. They lack a connection to an overall pedagogical plan that outlines what needs to be learned and motivates its relevance.

2.2 Task-Supported Language Teaching (TSLT)

The emphasis on functional learning goals is part of the foreign language didactic approach of task orientation, TSLT or TBLT (Ellis, 2003). While TBLT focusses on authentic language use from the beginning and incorporate explicit grammatical practice only incidentally, TSLT includes explicit grammar practice prior to meaning-oriented language use and is closer to school reality in Germany. Task-orientation is firmly established in the German curricula for English as a foreign language (Müller-Hartmann & Schocker von Ditfurth, 2011).

Communicative tasks, usually performed in class, form the centre of learning, instruction, and assessment (Spada, 2021). Ellis (2003) compiled a list of definitions for "task" in the foreign language context, which stress the task as an activity designed to functionally reach an objective. Since functionally successful language use is generally seen as the ultimate goal of foreign language learning, tasks sometimes are also referred to as "target tasks" that students should be able to perform (Long, 1985).

To be able to solve the target tasks, a "pre-task" phase can help students practice the task-essential language, i.e., the relevant language means (e.g., conditional clauses), while remaining meaning-focused (Ellis, 2003). In TSLT, this phase prior to the target tasks includes explicit form-focused exercises.

2.3 Learner Models and Task Orientation in Learner Dashboard Design

Building on the increasing attention to the interplay of technology, learners, and pedagogy (Lai & Li, 2011), it would be beneficial to incorporate more learner-centred *design for learning* principles into CALL research and practice, also adapting to the demands of classroom settings that are ecologically valid (Sun, 2017). OLMs provide such learner-oriented means of supporting learners to actively shape their own learning and pursue functional goals. Yet, OLMs aimed at infering the students' knowledge about the subject domain from the interaction data generally do not split the gathered information into subsets that are relevant for particular communicative functional language use.

For example, a student answer to an exercise that is part of the pre-task phase of a certain task may contain errors regarding multiple concepts (e.g., *irregular past tense verb form* and *conditional clauses type 2*). Some of these may be the pedagogical focus in a different target task, but information about the student performance on these concepts is updated in the learner model independently of the target task. The performance scores of concepts that are not the current focus of instruction thus may change in ways not transparent to the learner (Mabbott & Bull, 2004).

While OLMs provide relevant information for the learner, they were not designed to relate the learner's state and development of knowledge to a pedagogical learning goal. So, they do not indicate progress towards target tasks as part of a foreign language school curriculum as envisaged under a TSLT approach. Establishing this link also facilitates the effective implementation of ICALL systems in real-life school contexts since it then transparently supports TSLT as the established didactic approach. We therefore propose to combine OLM with the TSLT perspective by developing a transparently structured learner dashboard that provide insights on a student's learning status and progress towards goals related to target tasks.

To design a learner dashboard that combines OLM functionality with the TSLT perspective, we need to spell out the components of such a *student*, *learner*, or *learning* dashboard. For Schwendimann et al. (2017), "a learning dashboard is a single display that aggregates different indicators about learner(s), learning process(es), and/or learning context(s) into one or multiple visualisations" (p. 37). The relevant information can be derived from click-stream behavioural data, meta-information of the learning material, learner models, or a combination of these sources. According to Verbert et al. (2013), student dashboards should (1) make learners aware of their current learning status, (2) raise questions concerning the relevance of the displayed information for one's own learning, (3) provoke new insights, and (4) impact the learners' following learning behaviour. These stages take effect if the single display of a learner dashboard includes an explicit learning goal and valid, aggregated information on students' progress towards this learning goal or target task. We will argue that this can be achieved if the learner dashboard contains criterion-referenced feedback (Brown & Hudson, 2002).

What is criterion-referenced feedback? A target task or learning goal requires certain linguistic or communicative competencies, basically pre-established criteria, to be fulfilled by the learners (Mirmakhmudova, 2021). Although the term "criterion-referenced" appears more frequently with "assessment" or "testing," where performance is defined by achieving pre-defined criteria and not in relation to other students' performances (Lynch & Davidson, 1994), we here emphasise the value of this criterion-referenced measurement in a practice context. Setting acquisition criteria is a common approach in SLA: Learning a language is often regarded as a gradual development, which means that certain linguistic structures need to be emerged before others can be acquired (Pallotti, 2007). Criterion-referenced feedback considers learners' individual performance to these criteria (González-Marcos et al., 2019). For example, to write a report, learners need to be able to use the simple past and build regular and irregular verb forms. This is in accordance with TSLT's emphasis on the complex target tasks' communicative demands, which form the centre of this task-oriented teaching and learning (Robinson, 2011). Criterion-referenced feedback enables learners to obtain reliable information on their progress towards pre-defined goals and, thus, their ability to solve the target tasks (Mirmakhmudova, 2021). The success of including criterion-referenced feedback in students' learning has already been shown in analogue learning environments: Criterion-referenced feedback was revealed to be more effective on students' task performance than individual- or social-referenced feedback (Wilbert et al., 2010; Wollenschläger et al., 2011). However, research on criterion-referenced feedback and learner dashboards has so far been mainly conducted in higher education contexts (Schwendimann et al., 2017).

In the following, we will present a concrete approach to include criterion-referenced feedback in the context of task-supported learning in secondary school classrooms. For this, we enriched a practice-oriented ICALL system for English as a foreign language with a learner dashboard depicting students' progress towards pre-defined criteria and target tasks.

3 DEVELOPMENT OF A CRITERION-REFERENCED LEARNER DASHBOARD

We, a team of system developers, educational researchers, and SLA experts supported by English teachers, developed a so-called criterion-referenced learner dashboard that serves multiple functionalities to support students' language learning and metacognitive *learning to learn* (European Schoolnet, 2014; Fredriksson & Hoskins, 2008). On the one hand, the dashboard serves as a performance and progress view that gives students insights into their current learning status with respect to the criteria for the target task. On the other hand, it offers the navigation and selection of exercises, including additional practice opportunities to further progress towards the mastery of the language means needed for the target task. The dashboard's clear connection to the communicative target task is realised by listing the required grammatical and lexical criteria to fulfil the functional communicative goal stated in the header. This strengthens the link between the function and the form-focused practice.

Tailored to support the German school reality, the system is designed for TSLT classrooms, and best integrated into the pre-task practice phase, where language means are taught explicitly, while maintaining the functional goal in mind. In TBLT classrooms, our system best fits into the post-task phase where remaining knowledge gaps are being explicitly addressed with additional practice.

The dashboard is embedded in the existing ILTS FeedBook (Meurers et al., 2019), which provides immediate, scaffolded feedback on learners' responses. Scaffolded feedback provides incremental hints to enable the students to get from an initially incorrect answer to the correct one (Finn & Metcalfe, 2010). Each exercise in the FeedBook consists of multiple items to be solved, for instance, each gap in a gap-filling exercise represents one item. The FeedBook provides scaffolded feedback on the students' response per item (Rudzewitz et al., 2018; Ziai et al., 2019). It has been shown to positively impact students' language proficiency (Meurers et al., 2019).

3.1 Task-Related Structure of Learning Content

In TSLT, the goal is to be successful in the functional target task. To prepare for a target task, students practice the linguistic components needed with a given set of exercises that, together with the target task, form a task cycle[1]. Our goal was to integrate this approach into the digital tool. Therefore, didactic experts designed a variety of exercises for the FeedBook to explicitly foster students' acquisition of linguistic competencies required for the target task (Schmidt & Strasser, 2022). The digital exercises are thus intended to augment, not replace the classroom-based instruction. The content of the task cycles and thus the practice in the FeedBook is aligned with the German seventh-grade curriculum for English as a foreign language in academic-track schools.

For the development of our dashboard, we completed a few structural steps concerning the learning material. Following the TSLT approach, we defined criteria (i.e., grammatical and lexical language means that students must master to solve or successfully participate in the cycles' specific target task), which correspond to entities in the systems inherent learner model, and structured the exercises accordingly. Each language mean can be practiced in several exercises. To associate an exercise with a language mean, the main pedagogical objective of the exercise has been annotated manually by the didactic experts. In conclusion, the visual layout of the exercise structure is aligned with the internal learner model and thus the disconnect between practice and learner model decreases.

Based on this idea of embedded practice in task cycles, we developed the dashboard so that each cycle has its own corresponding dashboard view. The dashboard header (see Figure 1) provides the task orientation description by stating the target task title and the communicative goal. This header ensures that the task and, thus, the functional connection of the exercises, is always present and visible to the user. The dashboard can thus help to remind the students what they practice for during individual working times, such as homework – when no one else is there to motivate them.

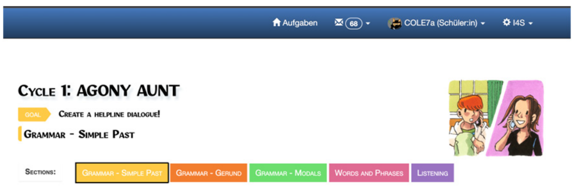

Figure 1: Task-oriented header for cycle 1, including the goal of the communicative target task, the different topics (called "sections").

Since each target task requires mastery of certain language means, these means are listed in the dashboard, and their associated exercises can be accessed via a click on the button labelled "practice" (see Figure 3). Students can work on these exercises in any order. Fine-grained language means (e.g., "gerunds after prepositions," "gerunds as subjects," or "gerunds as objects") are clustered into more coarse-grained grammatical topics (e.g., "grammar – gerund") and lexical topics (e.g., "words and phrases"). All topics of a cycle are accessible via the header (see Figure 1 and 3 as an example for Cycle 1). Only the language means of one topic are displayed at a time. This user-friendly structuring helps to reduce overloading the screen and therefore reduces scrolling down a list of 20 language means on one page, which can become tedious on small screen devices such as smartphones (Trewin, 2006).

3.2 Criterion-Referenced Performance Feedback

Besides the task-oriented header and the listing of relevant language means, each topic's dashboard view contains multiple components that together compose the learner dashboard.

The criterion-referenced performance feedback bar provides criterion-referenced information on students' performance and operates on the exercise level (see Figure 2 and 3). Performance bars are a typical game design element having the potential to satisfy learners' need for competence and to communicate the meaningfulness of the exercise (Sailer et al., 2017). Performance feedback on the exercise level is represented in an easily understandable horizontal stacked bar chart consisting of three parts (see Figure 2), each stating the absolute number of exercise items that fall into the category:

- *Correct at first try* (dark turquoise): If the student's initial response for an item matches the target answer (or one of multiple target answers).
- *Correct after feedback* (light turquoise): If the initial response for an item was incorrect but the student managed to get to the correct answer with the help of the scaffolded feedback provided via the system, regardless of the number of attempts.
- *Incorrect or missing* (grey): If the student submits an incorrect response or leaves an item unanswered.

[1] An example of a target task for which the system provides digital practice exercises can be found in Pili-Moss et al. (2024).

The number of hints received per individual item is not considered for the three categories, as taking feedback hints should not be seen as a punishment. The performance category description is available via the legend panel in the header (see Figure 3). Such performance bars provide a summary of the student's interaction with the exercise and count towards the acquisition of the respective language mean – and the latter is the ultimate criterion to be mastered for a successful participation in the target task. This holistic performance measurement on the exercise level is the foundation for the aggregated view in the criterion-referenced dashboard.

Figure 2: Exercise level performance stacked bar chart, consisting of three parts. One bar is mapped to one exercise; the different performance categories are colour coded. Dark turquoise for "correct at first try," light turquoise for "correct after feedback," and light grey for "incorrect or missing." The numbers in the circles represent the absolute number of items from the exercise falling into the respective performance category.

This aggregated dashboard view in Figure 3 visualises the learners' criterion-referenced performance on the language mean level, namely the completion accuracy. Per language mean, the exercises are ordered in the dashboard by increasing difficulty, from the least to the most complex. The difficulty level inherent to the exercise has been defined on a global level for all students by the didactic creators of the activity. Within this aggregated view, the individual performance bars are all scaled to the same size to be easily comparable.

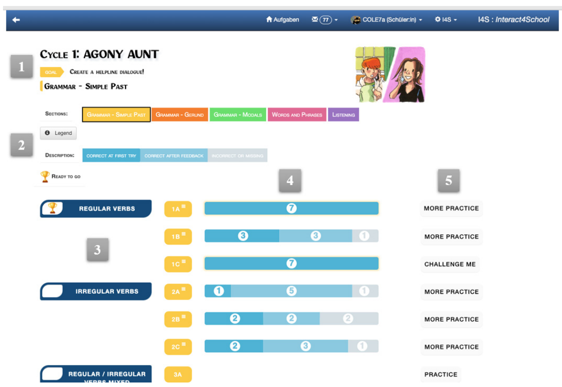

Figure 3: Criterion-referenced dashboard view for Cycle 1 – grammar topic "simple past," including the following components: (1) header, (2) expanded legend, (3) language means, (4) criterion-referenced performances, and (5) "more practice" functionality.

3.2.1 Design Procedure of the Criterion-Referenced Performance Bar

The development and design of the criterion-referenced visualisation and the final dashboard was an iterative process.

The idea of the stacked bar charts was based on existing dashboards, OLM visualisations, as well as learning games. However, none of the existing visualisations fit our purpose entirely, namely being suitable for seventh graders, complying with the exercise types and feedback format in the FeedBook, and enabling task orientation. The dashboard design process contained the involvement of the target group and the implementation of current scientific design principles (e.g., Bennett & Folley, 2019; Bodily et al., 2018; Bodily & Verbert, 2017; Sedrakyan et al., 2020).

To ensure comprehensibility for the users, we conducted two empirical studies. First, we ran an explorative user study. In this study, we interviewed six seventh-grade students using a "think-aloud" method and a guided interview to obtain qualitative insights into their understanding of the planned components. During the think-aloud part, students watched a screencast showing a fictional student using the FeedBook and the current dashboard. They needed to describe what was happening and interpret what they could see. During the subsequent interview, we showed them certain features again and asked them explicitly to explain their meaning. We presented the stacked bar chart accompanied by stars. These stars tried to map the criterion-referenced performance, split in its three categories, onto one combined score. We also presented them, as a variation to the stacked bar chart, two distinct metrics. One represented the count of correct responses at submission time as absolute numbers and one represented the percentage of successfully used hints (i.e., hints that resulted in uptake, and thus, correct solution). These separate metrics were less understandable and thus, we opted for the combined criterion-referenced performance bar. Second, 36 sixth-, seventh-, and eighth-grade students participated in an online survey with which we evaluated their understanding, preference, and the components' adequacy. According to the results, the criterion-referenced bar chart as a performance metric was easy to understand for the target group. We did not add stars to the bar chart for the final implementation, because we could not see a surplus value. The mapping was not trivial in terms of weighting the different criterion-referenced categories regarding the combined score, which was also reflected in the students' ratings, where they were required to assign a star score to a particular bar chart. The star scores (one to three stars, allowing for half stars) that the students set differed hugely between students and from the automatically calculated system score. Moreover, the system lacks a functionality that enables students to utilise collected stars as a form of currency for making purchases, which consequently may cause frustration. By implementing these decisions, the dashboard's complexity is reduced as it only needs to display a single combined metric instead of two.

For the bar chart itself, we decided against signal colours such as red to avoid evoking strong negative emotions. In addition, to accommodate for colour blindness, we decided against a distinct colour per criterion-referenced category, and instead opted for the bar chart ranging from a dark turquoise to a light grey, with the most intensive, brightest colour representing the target category "correct at first try." The bar makes the distinction between "correct at first try" and "correct after feedback" salient to the learner. It could technically be generalised over all exercise types the system provides (e.g., gap filling, multiple choice, jumbled sentence) and is thus a consistent measurement.

3.3 "More Practice" Opportunities – Distinction in Core and Parallel Exercises

In addition to the difficulty level, each exercise is manually annotated as either *core exercise* or *parallel exercise*. Core exercises are mandatory exercises. Each parallel exercise, in the interface labelled as *more practice*, is similar to its corresponding core exercise in terms of targeted competence, exercise type, and exercise difficulty but may differ in terms of vocabulary and syntax. Parallel exercises were designed to create enough practice opportunities to be able to reach the competence level needed to fulfil the goal and master the target task. These exercises thus enable students to improve their performance for a language mean at a particular difficulty level. Students can develop from system supported, scaffolded success (i.e., "correct after feedback") to independent success (i.e., "correct at first try"). Core exercises are directly accessible via the dashboard, and parallel exercises are accessible via the "more practice" button after their respective core exercise has been completed. Usually, the system offers multiple parallel exercises per core exercise (i.e., there are several core exercises per language mean, and one core exercise per difficulty level). The aggregated dashboard view (see Figure 3) displays one criterion-referenced performance per difficulty level per language mean. If a student has worked on multiple parallel exercises, only the best submission for this difficulty level is displayed. Depicting only their best submissions per difficulty level in the standard view might help students to focus on their progress and successes. By being able to hide less successful attempts, they might be more confident to show their dashboard view to others, including their parents or teachers. Moreover, learners can unfold the aggregated view by clicking the yellow exercise button to display all submissions for the language mean difficulty level (see Figure 4). Through this unfolding of the submission history, learners can see the submission of the core and parallel exercises ordered by submission date. The submission history can therefore be seen as an option to gain insight into the temporal progress with respect to the specific criterion (i.e., the language mean).

If students click on an exercise button in this history, they can decide to either look at their exercise submission again or re-open the exercise in practice mode (without their previous answers).

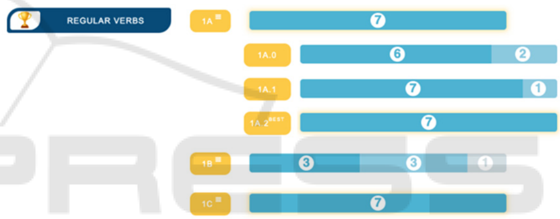

Figure 4: Language means exercise difficulty progression history unfolded for exercise 1A. The trophy on the top left indicates mastery of the language mean "regular verbs."

3.4 Language Mean Mastery – Proceeding to the Target Task

We have taken an additional step by incorporating a mastery measurement called "ready-to-go-ness." It is visualised as a small trophy icon and shows the students whether their performance on a language mean reached a certain level indicating if they are ready to proceed to the target task (see Figure 4). The idea to use a trophy, a badge, is derived from the gamification literature: Badges, similar to performance bars, have the potential to foster motivation and engagement by emphasising learners' competence (Hamari, 2017; Sailer et al., 2017). Gamified elements have also been the subject of research in CALL systems and have been reported to positively impact learners' learning experiences and outcomes (Dehghanzadeh et al., 2021). Goal setting can increase students' engagement and motivation (He & Loewen, 2022), which in turn can enhance the pedagogical effectiveness of digital learning environments (Bodnar et al., 2016).

The learner's intention to earn a badge (i.e., mastery trophy) can be regarded as such a goal. To prevent negative effects of badges, they are only visible to the students themselves, not peers (see Kyewski & Krämer, 2018).

What is required of students to earn a trophy? Accuracy of a language mean alone might not be a sufficient indicator of mastery (Pallotti, 2007). Therefore, we decided that ready-to-go-ness for a language mean comprises two components: accuracy and effort. To address the effort component of ready-to-go-ness, students must work on and submit all exercise difficulty levels in the language means cluster. Thus, exercises which have only been opened and submitted without student answers do not count towards meeting the effort requirement.

To address accuracy, students must reach a performance threshold at the diagnostic exercise of the language mean. The diagnostic exercise is the exercise with the difficulty level that is required to solve the respective target task. In our dashboard, it is always the bottom exercise of each language mean. Students earn a trophy, and thus, are "ready to go," if they solve at least 60% of the items "correct at first try" on the diagnostic core exercise or one of its parallel exercises. We defined mastery exclusively as "correct at first try" because this resembles an exam or test situation. Our decision to settle on this fixed threshold has multiple reasons. For this first version of ready-to-go-ness, we focused on introducing a measurement that is consistent throughout the whole system, works on all exercise types, and, thus, is transparent to the students and teachers. In a later stage, one could imagine setting a data-driven threshold per exercise or empirically refining the hardcoded threshold. However, submission data on the exercises is a prerequisite for both solutions. The definition of mastery can also be accessed by the users via the legend in the header. The earlier mentioned parallel exercises enable learners to improve and reach the required thresholds for a ready-to-go-ness trophy. The mastery criterion ready-to-go-ness is applied to each language mean.

Additionally, if students have achieved 100% "correct at first try" on a difficulty level (e.g., exercise 1C in Figure 4, the student has seven out of seven items "correct at first try"), the phrasing on the button changes from "more practice" to "challenge me." The different wording indicates to high-performing students that they have achieved an adequate level of proficiency in this language mean. However, it also provides additional opportunities for practice to further improve or showcase their skills.

The criterion-referenced performance bars, the mastery criterion visualised as trophy, and the "challenge me" phrasing can help students identify their strengths and weaknesses in relation to the target task. Thus, the dashboard presumably fosters students' metacognitive learning to learn.

4 CONCLUSION AND OUTLOOK

Implementing a learner dashboard including criterion-referenced feedback in an ILTS showed that practice-oriented ICALL and task orientation can be combined in a formal educational framework. Exercises to practice specific language means can be presented meaningfully, highlighting their connection to the target tasks and, thus, communicative learning goals. Our approach can be adapted to a variety of systems dealing with language learning, as language means could, for example, also correspond to "Can-Do-Statements" of the Common European Framework of Reference for Languages (CEFR) standard (Council of Europe, 2022).

A close collaboration between system developers, educational researchers, SLA experts, English didactics and practitioners was necessary so that, finally, the learner dashboard could meet the high standards of SLA in formal secondary education. The embedding of exercises in a meaningful context (here by structuring the ILTS based on task cycles) and other feedback features (such as the criterion-referenced stacked bar chart or trophy) described in this article, can also be adapted to learner dashboards within ILTS for other school subjects such as math.

Keeping the complexity of a learner dashboard at a minimal level while considering didactic prerequisites and user experience has been one of the major challenges for our endeavour. We aimed to find a balance between showing learners all insights they need to improve, and not overloading them with information. A limitation of the learner dashboard is the rather complicated definition of ready-to-go-ness. Students might struggle, for example, to understand how to earn a trophy and what it means to earn a trophy in relation to the target task. Another limitation might be that the aggregated dashboard view only depicts the best-performed exercise per language mean and difficulty, which does not necessarily reveal students' actual development. A future version of the dashboard could entail a single, but more elaborate, combined criterion-referenced performance score that considers all recent core and parallel exercise submissions. A "forgetting mechanism" considering only most recent submissions should be included in such a performance score. Such a mechanism would require many parallel exercises but allow students to "polish" their dashboard through more practice. We introduced a dual-purpose dashboard that combines exercise navigation and selection with an expandable task-oriented progress and performance view. In the dashboard, we visualised the pre-task activities, organised by task requirements (the language means), as stepping stones towards successfully fulfilling the target task. Therefore, we introduced three new concepts:

- An easy-to-understand criterion-referenced exercise performance feedback (including "correct at first try")
- Parallel exercises for further practice
- Operationalisation of acquisition criteria via mastery criterion (ready-to-go-ness)

Realising this prototype of a criterion-referenced learner dashboard is only the start for further research on combining task orientation and ICALL. The availability of a task-supported learning ILTS for seventh-grade English learners evokes the question of how its actual field use impacts students' learning and in-class participation: Does the additional, intensive pre-task phase when students practice with the FeedBook help students to solve the in-class target task more successfully? Does the criterion-referenced feedback influence students' motivation or perceived competence? And ultimately, is the criterion-referenced dashboard beneficial for students' language acquisition and proficiency?

To answer these questions, the FeedBook, including the criterion-referenced learner dashboard, was used in a multi-site cluster-randomised controlled field trial in the context of the Interact4School project (see Parrisius et al., 2022). Among other variations, classes participating in the study received either access to the entire criterion-referenced learner dashboard or a control version without the performance bar when working with the FeedBook. The use of the ILTS and dashboard throughout the whole school year was accompanied by several data collections focusing on assessing students' current language competencies and motivation. First investigations showed that students performed better in task cycle-specific performance tests if they had access to the learner dashboard compared with students who had no access (Parrisius, Wendebourg, Rieger, et al., 2024). Further, if students reported high initial motivation for English, they experienced positive effects of the learner dashboard on their subsequent motivation (Parrisius, Wendebourg, Pieronczyk, et al., 2024).

Other follow-up projects will focus on mainly two further aspects. First, to address heterogeneity, the task-oriented dashboard will be extended and adapted for adaptive exercise sequencing. Second, so far, teachers have limited access to the performance-based information the students receive through the dashboard. In a next step, we will examine how to best aggregate the learners' information on a class level in the form of a teacher dashboard and how such a dashboard impacts teaching decisions.

ACKNOWLEDGEMENTS

We thank Kristina Dawidowsky from the University of Tübingen for her expertise and contribution in the field of user interface design and user experience. We also thank Carolyn Blume from the TU Dortmund, Diana Pili-Moss, and Torben Schmidt from the Leuphana University Lüneburg for their project cooperation and creation of exercises for the FeedBook. This research is based on work carried out in the project Interact4School which is funded by the German Federal Ministry of Education and Research (BMBF 01JD1905A & 01JD1905B). The authors are responsible for the content of this publication.

REFERENCES

Bennett, L., & Folley, S. (2019). Four design principles for learner dashboards that support student agency and empowerment. *Journal of Applied Research in Higher Education*, *12*(1), 15–26.

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakay, F., & Verbert, K. (2018). Open learner models and learning analytics dashboards: A systematic review. *LAK'18: International Conference on Learning Analytics and Knowledge, March 7-9*, 41–50. https://doi.org/10.1145/3170358.317040

Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, *10*(4), 405–418. https://doi.org/10.1109/TLT.2017.2740172

Bodnar, S., Cucchiarini, C., Strik, H., & van Hout, R. (2016). Evaluating the motivational impact of CALL systems: current practices and future directions. *Computer Assisted Language Learning*, *29*(1), 186–212. https://doi.org/10.1080/09588221.2014.927365

Brauner, U., & Prediger, S. (2018). Alltagsintegrierte Sprachbildung im Fachunterricht - Fordern und unterstützen fachbezogener Sprachhandlungen. In C. Titz, S. Geyer, A. Ropeter, H. Wagner, S. Weber, & M. Hasselhorn (Eds.), *Konzepte zur Sprach- und Schriftsprachförderung entwickeln* (pp. 228–248). Kohlhammer.

Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge University Press. https://doi.org/9780521000833

Brusilovsky, P. & Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web* (Vol. 4321, pp. 3–53). Springer. https://doi.org/10.1007/978-3-540-72079-9_1

Bull, S. (2004). Supporting learning with open learner models. *Planning*, *29*(14).

Bull, S., Ginon, B., Boscolo, C., & Johnson, M. (2016). Introduction of learning visualisations and metacognitive support in a persuadable open learner model. *LAK '16: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 30–39. https://doi.org/10.1145/2883851.2883853

Bull, S., & Kay, J. (2010). Open learner models. *Studies in Computational Intelligence*, *17*(308), 301–322. https://doi.org/10.1007/978-3-642-14363-2

Bull, S., & Pain, H. (1995). Did I say what I think I said, and do you agree with me? In *Inspecting and Questioning the Student Model* (pp. 501–508). University of Edinburgh, Department of Artificial Intelligence.

Bull, S., & Wasson, B. (2016). Competence visualisation: Making sense of data from 21st-century technologies in language learning. *ReCALL*, *28*(2), 147–165. https://doi.org/10.1017/S0958344015000282

Canals, L., & Mor, Y. (2020). Towards a signature pedagogy for task-based technology-enhanced language learning: Design patterns. *ReCALL*, *35*(1), 4–18. https://doi.org/10.1145/3424771.3424787

Chen, X. L., Zou, D., Xie, H. R., & Su, F. (2020). Twenty-five years of computer-assisted language learning: A topic modeling analysis. *Language Learning & Technology*, *25*(3), 151–185. http://hdl.handle.net/10125/73454

Chong, S. W., & Reinders, H. (2020). Technology-mediated task-based language teaching: A qualitative research synthesis. *Language Learning & Technology*, *24*(3), 70–86. https://research-repository.st-andrews.ac.uk/handle/10023/24982

Council of Europe. (2022). *Common European framework of reference for languages: Learning, teaching, assessment (CEFR)*. https://coe.int/en/web/common-european-framework-reference-languages/home

Davies, G., Otto, S. E., & Rüschoff, B. (2012). Historical perspectives on CALL. In Thomas, M., Reinders, H. & Warschauer, M. (eds.), *Contemporary computer-assisted language learning*. New York, NY: Bloomsbury, 19-38.

Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaee, E., & Noroozi, O. (2021). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning*, *34*(7), 934–957. https://doi.org/10.1080/09588221.2019.1648298

DeKeyser, R. (2010). Practice for second language learning: Don't throw out the baby with the bathwater. *IJES*, *10*(1), 155–165. https://um.es/ijes

DeKeyser, R. (2020). Skill acquisition theory. In B. VanPatten, D. Keating, Gregory, & S. Wulff (Eds.), *Theories in Second Language Acquisition. An Introduction* (3rd ed., pp. 95–112). Routledge.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

European Schoolnet. (2014). *Learning to learn*. KeyCoNet. http://keyconet.eun.org/learning-to-learn

Finn, B., & Metcalfe, J. (2010). Scaffolding feedback to maximize long-term error correction. *Memory and Cognition*, *38*(7), 951–961. https://doi.org/10.3758/MC.38.7.951

Fredriksson, U., & Hoskins, B. (2008). Learning to learn: What is it and can it be measured? In *EUR 23432 EN. Luxembourg (Luxembourg): OPOCE*. JRC. https://doi.org/10.2788/83908

González-Marcos, A., Navaridas-Nalda, F., Ordieres-Meré, J., & Alba-Elías, F. (2019). A model for competence e-assessment and feedback in higher education. In A. Azevedo & J. Azevedo (Eds.), *Handbook of Research on E-Assessment in Higher Education* (pp. 295–311). IGI Global. https://doi.org/10.4018/978-1-5225-5936-8.ch012

Hamari, J. (2017). Do badges increase user activity? A field experiment on the effects of gamification. *Computers in Human Behavior*, *71*, 469–478. https://doi.org/10.1016/j.chb.2015.03.036

He, X., & Loewen, S. (2022). Stimulating learner engagement in app-based L2 vocabulary self-study: Goals and feedback for effective L2 pedagogy. *System*, *105*, 102719. https://doi.org/10.1016/j.system.2021.102719

Heift, T., & Schulze, M. (2003). Student modeling and ab initio language learning. *System, 31*(4), 519-535.

Kolb, A., & Raith, T. (2018). Principles and methods—Focus on learners, content and tasks. *Teaching English as a foreign language: An introduction,* 195-209.

Kos, T. (2023). A teacher-researcher snapshot of task-based peer interactions in EFL secondary school classrooms in Germany. *Language Teaching for Young Learners, 5(2)*, 170-195.

Kyewski, E., & Krämer, N. C. (2018). To gamify or not to gamify? An experimental field study of the influence of badges on motivation, activity, and performance in an online learning course. *Computers & Education*, *118*, 25–37. https://doi.org/10.1016/j.compedu.2017.11.006

Lai, C., & Li, G. (2011). Technology and task-based language teaching: A critical review. *CALICO Journal*, *28*(2), 498–521. http://jstor.org/stable/calicojournal.28.2.498

Long, M. H. (1985). A role for instruction in second language acquisition: Task-based language teaching. In K. Hyltenstam & M. Pienemann (Eds.), *Modelling and Assessing Second Language Acquisition* (pp. 77–99). Multilingual Matters.

Long, M. H. (1991). Focus on form: A design feature in language teaching methodology. *Foreign Language Research in Cross-Cultural Perspective*, 39–52.

Long, Y., & Aleven, V. (2017). Enhancing learning outcomes through self-regulated learning support with an Open Learner Model. *User Modeling and User-Adapted Interaction*, *27*(1), 55–88. https://doi.org/10.1007/S11257-016-9186-6

Lynch, B. K., & Davidson, F. (1994). Criterion-referenced language test development: Linking curricula, teachers, and tests. *Quarterly*, *28*(4), 727–743.

Mabbott, A., & Bull, S. (2004). Alternative views on knowledge: Presentation of open learner models. *International Conference on Intelligent Tutoring Systems*, *3220*, 689–698. https://doi.org/10.1007/978-3-540-30139-4_65

Meurers, D., De Kuthy, K., Nuxoll, F., Rudzewitz, B., & Ziai, R. (2019). Scaling up intervention studies to investigate real-life foreign language learning in school. *Annual Review of Applied Linguistics*, *39*(2019), 161–188. https://doi.org/10.1017/S0267190519000126

Mirmakhmudova, I. I. (2021). Comparing criterion and norm referenced assessments of langauge skills in the second language. *Asian Journal of Social Sciences & Humanities*, *11*(11). https://doi.org/10.5958/2249-7315.2021.00242.2

Molenaar, I., & Knoop-Van Campen, C. (2016). Learning analytics in practice: The effects of adaptive educational technology Snappet on students' arithmetic skills. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 538–539. https://doi.org/10.1145/2883851

Müller-Hartmann, A., & Schocker von Ditfurth, M. (2011). *Teaching English: Task-supported language learning*. Schöningh.

Pallotti, G. (2007). An operational definition of the emergence criterion. *Applied Linguistics*, *28*(3), 361–382. https://doi.org/10.1093/applin/amm018

Parrisius, C., Pieronczyk, I., Blume, C., Wendebourg, K., Pili-Moss, D., Assmann, M., Beilharz, S., Bodnar, S., Colling, L., Holz, H., Middelanis, L., Nuxoll, F., Schmidt-Peterson, J., Meurers, D., Nagengast, B., Schmidt, T., & Trautwein, U. (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): Pre-registration of the study design*. PsychArchives. https://doi.org/10.23668/psycharchives.5366

Parrisius, C., Wendebourg, K., Pieronczyk, I., Holz, H., Deininger, H., Schmidt, T., Meurers, D., Nagengast, B., & Trautwein, U. (2024*). Examining effects of gamification elements in an intelligent tutoring system for 7th grade English learners on their motivation – A randomized controlled field trial* [Unpublished manuscript]. Karlsruhe University of Education, Germany.

Parrisius, C., Wendebourg, K., Rieger, S., Blume, C., Pili-Moss, D., Colling, L., Pieronczyk, I., Holz, H., Bodnar, S., Loll., I., Schmidt, T., Trautwein, U., Meurers, D., & Nagengast, B. (2024). *Effective features of feedback in an intelligent language tutoring system* [Unpublished manuscript]. Karlsruhe University of Education, Germany.

Pili-Moss, D., Schmidt, T., Beilharz, S., & Blume, C. (2024). Projekt Interact for School (I4S). Unterrichtseinheit und Target-Task zum Thema "Giving Advice", Language Focus: The Past Tense, Gerunds, Modals; Grade 7. https://dio.org/10.48548/ pubdata-107

Ring, M., Brahm, T., & Randler, C. (2019). Do difficulty levels matter for graphical literacy? A performance assessment study with authentic graphs. *International Journal of Science Education*, *41*(13), 1787–1804. https://doi.org/10.1080/09500693.2019.1640915

Robinson, P. (2011). Task-based language learning: A review of issues. *Language Learning*, *61*, 1–36. https://doi.org/10.1111/j.1467-9922.2011.00641.x

Rudzewitz, B. (2021). *Learning analytics in intelligent computer-assisted language learning*. University of Tuebingen.

Rudzewitz, B., Ziai, R., De Kuthy, K., Möller, V., Nuxoll, F., & Meurers, D. (2018). *Generating feedback for English foreign language exercises*. 127–136. https://doi.org/10.18653/v1/w18-0513

Rudzewitz, B., Ziai, R., Nuxoll, F., De Kuthy, K., & Meurers, D. (2020). Enhancing a web-based language tutoring system with learning analytics. *CEUR Workshop Proceedings*, *2592*, 1–7.

Ruiz, S., Rebuschat, P., & Meurers, D. (2023). Individualization of practice through Intelligent CALL. In Y. Suzuki (Ed.), *Practice and automatization in second language research: Theory, methods, and pedagogical implications.*

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, *69*, 371–380. https://doi.org/10.1016/j.chb.2016.12.033

Schmidt, T., & Strasser, T. (2022). Artificial intelligence in foreign language learning and teaching: A CALL for intelligent practice. *Anglistik: International Journal of English Studies*, *33*(1), 165–184.

Schulze, M. & Heift, T. (2012). Intelligent CALL. In Thomas, M., Reinders, H. & Warschauer, M. (eds.), *Contemporary computer-assisted language learning*. New York, NY: Bloomsbury, 249–265.

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, *10*(1), 30–41. https://doi.org/10.1109/TLT.2016.2599522

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, *107*, 1–25. https://doi.org/10.1016/j.chb.2018.05.004

Spada, N. (2021). Reflecting on task-based language teaching from an instructed SLA perspective. *Language*

Teaching. https://doi.org/10.1017/S0261444821000161
Sun, S. Y. H. (2017). Design for CALL – possible synergies between CALL and design for learning. *Computer Assisted Language Learning*, *30*(6), 575–599. https://doi.org/10.1080/09588221.2017.1329216
Trewin, S. (2006). Physical usability and the mobile Web. *W4A '06: Proceedings of the 2006 International Cross-Disciplinary Workshop on Web Accessibility (W4A): Building the Mobile Web: Rediscovering Accessibility?*, 109–112. https://doi.org/10.1145/1133219.1133239
Tsourounis, S., & Demmans Epp, C. (2016). Learning dashboards and gamification in MALL: Design guidelines in practice. In A. Palalas & M. Ally (Eds.), *The International Handbook of Mobile-Assisted Language Learning* (pp. 370–398). China Central Radio & TV University Press Co., Ltd.
Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, *57*(10), 1500–1509. https://doi.org/10.1177/0002764213479363
Vogt, K., & Schmidt, T. (2021). Digitale Transformation im Fremdsprachenunterricht und dessen Bildungsauftrag. In C. Maurer, K. Rincke, & M. Hemmer (Eds.), *Fachliche Bildung und digitale Transformation - Fachdidaktische Forschung und Diskurse. Fachtagung der Gesellschaft für Fachdidaktik 2020* (pp. 44–47). https://doi.org/10.25656/01:21659
Wilbert, J., Grosche, M., & Gerdes, H. (2010). Effects of evaluative feedback on rate of learning and task motivation: An analogue experiment. *Learning Disabilities: A Contemporary Journal*, *8*(2), 43–52. https://researchgate.net/publication/208600646
Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford University Press.
Wollenschläger, M., Möller, J., & Harms, U. (2011). Effekte kompetenzieller Rückmeldung beim wissenschaftlichen Denken. *Zeitschrift Für Pädagogische Psychologie*, *25*(3), 197–202. https://doi.org/10.1024/1010-0652/a000040
Xu, J., & Bull, S. (2010). Encouraging advanced second language speakers to recognise their language difficulties: A personalised computer-based approach. *Computer Assisted Language Learning*, *23*(2), 111–127. https://doi.org/10.1080/09588221003666206
Ziai, R., Nuxoll, F., Kuthy, K. De, Rudzewitz, B., & Meurers, D. (2019). The impact of spelling correction and task context on short answer assessment for intelligent tutoring systems. *Proceedings of the 8th Workshop on NLP for Computer Assisted Language Learning*, 93–99. https://aclweb.org/anthology/W19-6310