

ARTICLE



## Learning processes in interactive CALL systems: Linking automatic feedback, system logs, and learning outcomes

*Bronson Hui\**, University of Maryland, College Park

*Björn Rudzewitz\**, Department of Linguistics, University of Tübingen

*Detmar Meurers*, Department of Linguistics, University of Tübingen

*\*The first two authors contributed equally to this project.*

### Abstract

*Interactive digital tools increasingly used for language learning can provide detailed system logs (e.g., number of attempts, responses submitted), and thereby a window into the user's learning processes. To date, SLA researchers have made little use of such data to understand the relationships between learning conditions, processes, and outcomes. To fill this gap, we analyzed and interpreted detailed logs from an ICALL system used in a randomized controlled field study where 205 German learners of English in secondary school received either general or specific corrective feedback on grammar exercises. In addition to explicit pre-/post-test results, we derived 19 learning process variables from the system log. Exploratory factor analysis revealed three latent factors underlying these process variables: effort, accuracy focus, and time on task. Accuracy focus and finish time (a process variable that did not load well on any factors) significantly predicted pre-/post-test gain scores with a medium effect size. We then clustered learners based on their process patterns and found that the specific feedback group tended to demonstrate particular learning processes and that these patterns moderate the advantage of specific feedback. We discuss the implications of analyzing system logs for SLA, CALL, and education researchers and call for more collaboration.*

**Keywords:** *System Logs, Learning Processes, Feedback, Learner Clusters*

**Language(s) Learned in This Study:** *English*

**APA Citation:** Hui, B., Rudzewitz, B., & Meurers, D. (2023). Learning processes in interactive CALL systems: Linking automatic feedback, system logs, and learning outcomes. *Language Learning & Technology*, 27(1), 1–23. <https://hdl.handle.net/10125/73527>

### Introduction

An advantage of computer-assisted language learning (CALL) systems is that such systems can provide access to detailed logs of the learners' interaction. For example, one can log the number of attempts by a learner for a given item, the answers submitted in each attempt, and the system's feedback or other responses. While offering a potential wealth of information, such system logs need to be analyzed and interpreted carefully to become useful for stakeholders such as researchers, teachers, and system designers, to inform pedagogy and system designs (Gašević et al., 2015; Mangaroska & Giannakos, 2018; Siemens &

Baker, 2012). These fine-grained longitudinal records are particularly valuable for quantitative analyses of the users' learning processes. This line of investigation, placing emphasis upon learning processes, aligns with recent calls of second language acquisition (SLA) researchers for an increased awareness to distinguish language learning products (i.e., outcomes of learning as a result of treatment and/or teaching) and language learning processes (i.e., happening during treatment and/or teaching) and to lay a greater emphasis on the latter (e.g., Godfroid, 2019; Leow, 2015). In addition, log data allows researchers to identify learner clusters that are characterized by particular behavior patterns and the potential differential effects of certain interventions (e.g., provision of corrective feedback) based on such patterns.

To date, SLA researchers have not yet taken systematic advantage of CALL system logs to understand learning as a process despite the growing popularity of language tutoring systems. In this study, we took a first step to fill this gap by interpreting logs extracted from a system used throughout the entire school year in ten German secondary school classes of English as a foreign language. The primary goal of this paper is to spell out and illustrate how such logs can shed light on language learning processes, and crucially, on their association with learning conditions and outcomes. As secondary goals, we also examine the extent to which learner clusters can be identified based on learning process variables, and how they interacted with the provision of specific corrective feedback in accounting for eventual learning gains.

## Literature Review

While one goal of SLA research is to understand how learners acquire the target language, much work has focused on measuring learning products (e.g., learning gains on grammar tests), with less attention paid to the *process* of learning (e.g., Godfroid, 2019; Leow, 2015). Learning products represent outcomes of learning as a result of treatment and/or teaching, whereas learning processes are what happens *during* learning and teaching. Various learning processes, however, may or may not result in eventual learning as typically measured by pre-/post-tests. Therefore, studying learning products alone thus is not sufficient to understand important questions in SLA such as what processes are conducive to learning and ways to promote these processes. On this account, SLA researchers should study both learning products and processes.

To bring learning processes into the picture, researchers have championed the use of online methodologies such as eye tracking and verbal reports (Godfroid, 2019; Leow, 2015). These methodologies are particularly well-suited for measuring and quantifying attention and awareness as core SLA constructs (Robinson et al. 2013), to gather empirical evidence for (or against) theoretical claims about how second language learning takes place (i.e., the learning process). For example, Leow's (2015) model of explicit second language learning proposes three processing stages that result in learning: input, intake, and knowledge processing. With sufficient attention and awareness (input processing), linguistic input is noticed and becomes intake. Intake is further processed with cognitive effort (intake processing) to potentially be integrated with the existing knowledge system. During production, learners monitor their output (knowledge processing) in relation to their second language (L2) system. Focusing on the input processing stage, eye-tracking research on vocabulary learning, for example, has shown that longer reading times correlate positively with ultimate learning gains (e.g., Elgort et al., 2018; Godfroid et al., 2018). In a landmark study, Godfroid et al. (2013) embedded English-like pseudowords (as learning targets) into short passages for their Dutch learners of English to read. The authors found that learners who spent longer time on individual target words had a greater chance of recognizing them on a surprise vocabulary post-test. Subsequent studies using longer reading materials have also suggested that learners spend less time on unfamiliar words when they encounter the same lexical items repeatedly (Elgort et al., 2018; Godfroid et al., 2018). Although learners do not necessarily process the words in a more automatic manner (Hui, 2020), such speed ups may index a type of implicit and procedural learning of words (Godfroid, 2020). Importantly, with eye tracking, researchers can observe the stepwise learner behavior while the learning happens, supporting interpretation of the data in terms of the learning process. In sum, eye tracking represents a key online method that SLA researcher have relied on when studying language learning processes.

Similarly, Cerezo et al. (2016) presented evidence based on verbal reports that deeper processing can result in intake or even L2 knowledge. In the study, the authors provided learners with corrective feedback on the Spanish *gustar* constructions (i.e., an intransitive verb that takes a dative experiencer) in a video game. The feedback given to learners was meant to “gradually promote[s] deeper processing” (p. 274). The key question was then the extent to which this feedback could result in awareness of the knowledge being learned and hence eventual learning gains. Think-aloud data (verbalization of thoughts during learning) confirmed an association between deeper processing, awareness, and learning outcomes (as measured by a written production and a multiple-choice recognition activity). Another study by Cerezo (2016) similarly highlighted the importance of investigating language learning as a process. The author found counter-productive roles of fatigue and boredom in the learning process, accounting for a lack of association between amount of practice and learning products. In a study only measuring learning outcomes (i.e., without collecting and analyzing think-aloud data), the cause of such a lack of association would have been puzzling. Taken together, the studies showcase that examining evidence on the learning process is crucial for understanding learning.

While eye tracking and verbal reports are two key tools that allow SLA researchers to gain insights into learning processes, they are not without limitations. Think-aloud and eye-tracking data are time-consuming to collect and in practice cannot be collected for large-scale studies conducted outside of lab settings. In a digital learning context, system logs can offer complementary opportunities given that they can capture the individual learner behavior and performance in the system, which one can then try to interpret as a proxy for learning processes. By associating system log entries automatically recorded when users interact with the application with eventual learning outcomes, researchers can turn this wealth of information into meaningful interpretations of learning in real-life contexts in a way that can scale to a very large number of learners. While so far this line of research is only nascent in SLA, learning analytic approaches to interpreting learner logs are increasingly common in mathematics and the natural sciences. For example, Chen et al. (2018) investigated the relationship between system logs and eventual learning outcomes for 474 undergraduates taking an online physics module. The authors extracted a total of 30 variables (e.g., correct checks, intervals between checks, total session time) from the system logs and reported that 11 behavioral variables were important in accounting for up to 32% of the variance in course performance. However, only First to Due (i.e., the time between the first check and time it was due) emerged as a significant predictor of cumulative grades, while only First to Due and Average Session Length predicted final exam scores significantly. Further, the authors identified patterns of behavior that were associated with better grades. Cluster analysis indicated that higher performing learners tended to demonstrate self-regulation or productive struggles. For example, they tended to check their answers multiple times and to start and submit the homework early. With these findings, the authors argue that online learning platforms allow researchers to understand the process of learning “at a deeper level” (p. 73).

A rare example for such analyses in the SLA domain makes use of data collected by the E-Tutor system (Heift, 2010), a language tutoring system for German at the university level. The system provides meta-linguistic feedback to different types of activities and highlights errors in learner answers. Heift (2019) presents a longitudinal study linking uptake observations, as extracted from log data, to different learning conditions. Meta-linguistic feedback led to more uptake (correct answers after feedback) when compared with unspecific feedback providing only highlighting.

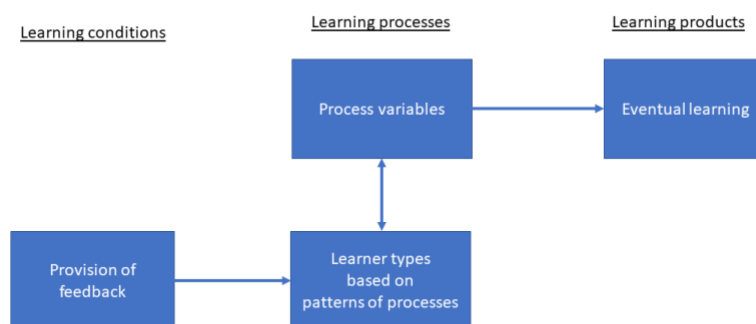
Summing up, while hardly used in SLA research so far, system logs extracted from CALL systems can provide information that in principle offer a window into the learning process as it unfolds. Both for basic research into the learning process and for applied work designed to advance systems and interventions, it thus is well-motivated to systematically explore using such log data, and to advance the methodology needed to validly interpret it to foster our understanding of the complex relationship between learning conditions, processes, and outcomes.

## The Present Study

The primary goal of the present study is to explore and refine the methodology needed to interpret system logs for SLA research into language learning processes. Using this window into learning processes, we investigated the relation between learning conditions, processes, and outcomes in the context of an intelligent computer-assisted learning (ICALL) system. First, we focused on the relationship between learning processes and products. We hypothesized that certain learning processes have an impact on learning products as measured by independent, explicit pre- and post-tests. Second, we focused on learning processes and were interested in identifying process patterns meaningfully distinguishing groups of learners. Finally, we examined how receiving specific feedback on grammar exercises may drive learning. We visualize these relationships in [Figure 1](#), illustrating how the learning conditions, processes, and products are linked, allowing feedback to influence learning process variables in a way that is mediated by learner types to lead to differential learning outcomes.

**Figure 1**

*Relationships Between Learning Conditions, Processes, and Products*



Turning hypotheses to specific research questions, we formulated the following questions to guide the present study:

1. RQ1: To what extent can learning process variables, as extracted from system logs, directly account for learning outcomes?
2. RQ2: Can we meaningfully distinguish clusters of learners based on the learning process variables?
3. RQ3: To what extent does specific feedback relate to the learning process clusters?

## Methodology

The current study analyzes data collected but not reported or examined by Meurers et al. (2019). With regard to the effectiveness of automated corrective feedback for grammar learning, the previous study examined the differences in learning outcomes between those who received specific feedback and those who received general feedback. In contrast, the current study places the emphasis on the learning processes as revealed by system logs. Below we provide an overview of the study context, the participants, and the instruments. Due to the focus of the present study, we refer readers to Rudzewitz et al. (2017, 2018) for a technical description of the system and here only provide the background information needed to understand the analysis and interpretation of the logs.

## Context of the Study and the System

The data were collected over an entire school year between September 2018 and July 2019 in authentic secondary school (“Gymnasium”) contexts involving five schools and 14 classes in the district of Tübingen in Baden-Württemberg, Germany. All students were 7th grade learners in their third year of formal English instruction in secondary school. Class sizes ranged from 15 to 31 (mean = 25). All schools were co-educational, but one class had only male students. The sampling included both public and private schools as well as a school that had an emphasis on science and one that focused on arts and sports. Three classes implemented a content and language integrated learning curriculum. Two were tablet classes, where students bring a tablet to school every day. Overall, this sample represented the diverse English education landscape in the area. After removing data for which no consent for scientific study use was given by the students and their parents, the final sample consisted of 205 students in ten classes.

The FeedBook system is a tutoring system for English as a foreign language. FeedBook is a web-based adaptation of a paper workbook, Camden Town Gymnasium 3, approved for use in local high schools. Hence, it models the complete curriculum. In total, there are 85 activities with different question formats (i.e., fill-in-the-blanks activities, short answer activities, and true/false exercises). The system was implemented as part of the regular teaching and learning routine. We did not constrain learners and teachers in how they used the system to maximize the ecological validity of the study. Teachers and students were free to use and interact with the system as they saw fit, in line with how the paper-based version of the workbook would have been used. The only requirements were that the teacher needed to assign a set of around five core activities for each chapter and that they needed to proceed through the chapters one by one. Learners had the opportunity to go beyond the activities assigned by their teacher and complete additional exercises for themselves. As learned from informal, personal conversations, teachers typically assigned exercises in the system as homework. The teachers sometimes used the learning analytics functions in the system to identify which errors were frequent/typical for a specific exercise, which became the basis of discussion in subsequent teaching. Face-to-face classroom teaching time was used primarily for communicative tasks, group work, and discussion of new grammar topics.

Although any language learning applications have the potential to log learners’ behavior, a system that supports human-computer interaction and automated provision of individualized feedback provides even richer information. FeedBook is an example in this regard. The main functionality for students is the immediate scaffolding feedback on their answers to grammar exercises (see descriptions below). FeedBook handles the variability in learner answers by abstracting away from concrete language forms, and instead, by specifying patterns that can be flexibly matched (based on a rank of priorities) against the learner’s production (Ziai et al., 2018). In particular, the system models common misconceptions that were generalized based on data collected with a previous version of the system (Rudzewitz et al., 2018). For each common misconception, an experienced German high school teacher specified a feedback message and the conditions of when it should be applied. At runtime, the system compares the input with the target answer. If the answer is not correct, the system searches for the best-matching alternative answer associated with a misconception before providing a feedback message. For students, the feedback messages pinpoint the exact source of misconceptions, and most importantly, how to correct the error. In the example shown in [Figure 2](#), the student typed *\*more higher* when *higher* was expected. The system recognized the potential misconception and provided the student with a strategy to correct the form.

## Figure 2

### Interactive Scaffolding Feedback

**B 6 Compare these people**  
Match the verbs and the adverbs and use them to compare these people.  
• LIF9 Comparison of adverbs

work · speak · drive · jump · sing · ride a bike ·  
hard · dangerously · high · well · fast · loudly

1

Cara / Rachel  
*Cara jumps more higher than Rachel.*

Feedback für "Cara jumps more higher than Rachel."  
When an adjective has one syllable, we use '-er/-est' for comparisons, not 'more/most'.  
Hilfreich?  
 Ja  Nein

Meurers et al. (2019) reported some initial findings of the implementation of FeedBook. Students in the intervention group receiving specific grammar feedback for their incorrect responses (see Figure 2 for an example) outperformed those in the control group who received general feedback on correctness (e.g., “This is not what I am expecting. Please try again.” [p. 175]). Regardless of feedback type, the learner often made another attempt which may lead to another feedback if the answer was still not correct.

### Product and Process Measures

In line with Meurers et al. (2019), we report data from chapter two of the workbook. The present data set included results of a grammar test (pre-/post-tests) and process variables extracted from system logs that captured the students’ performance and use of the system. In chapter two, a total of three grammatical constructions were covered: type-II conditionals, adverbs of comparison, and relative clauses.

### Product Measures

The result of this measure was reported in Meurers et al. (2019). There was a total of 40 items targeting the learner’s grammatical knowledge. They were written by an author of the same series as the paper version of the workbook. We analyzed two question types: grammaticality judgement items in the form of true/false activities (see Figure 3) and fill-in-the-blanks activities (see Figure 4). We obtained alpha reliabilities of .72 (95% CIs [.66, .77]) and .77 (95% CIs [.72, .81]) for the data obtained from the pre- and post-tests respectively. Students were able to navigate between the different items across question types and edit their answers freely. In the spirit of Open Science, these items have been made available on the IRIS repository (Marsden et al., 2016) under the publication by Meurers et al. (2019).

Figure 3

*Grammaticality Judgment Task*

Figure 4

*Fill-in-the-Blanks Task*
**Process Measures**

We operationalized the construct of learning process as learners' performance on and use of the system. To generate these quantitative variables, we first defined three critical time frames for analysis and variable generation. First, we defined a task as a collection of items on the same target construction (e.g., conditionals)—so we here use the term task not in the sense of task-based language learning but as used in the tutoring system literature in order to be able to connect consistently to established notions such as time on task. Concretely, task in this article denotes an exercise displayed on one page in the system, as adopted



from the paper version underlying the web application. In Figure 4, for example, there were ten items in the task. When a learner entered a particular task, in terms of timeframe, it was the start of a *task period (Level 1)*. The total time spent at this level, for example, captured the time on task in the learner's working on any of the ten items. The next, lower timeframe level was the task field level at which a learner worked on a particular blank that needed to be filled in. Note that one item could have more than one task field (e.g., an item having two blanks). The time spent on one task field was then a *task field period (Level 2)*. Finally, when a learner attempted a task field multiple times, they made multiple attempts in the system. Each *attempt period (Level 3)* was then defined from the last attempt to the current task field to the check of the present attempt. The system automatically logged off after 30 minutes of inactivity to avoid blocking access to the system for active users.

**Figure 5**

*Example of Time Frames*

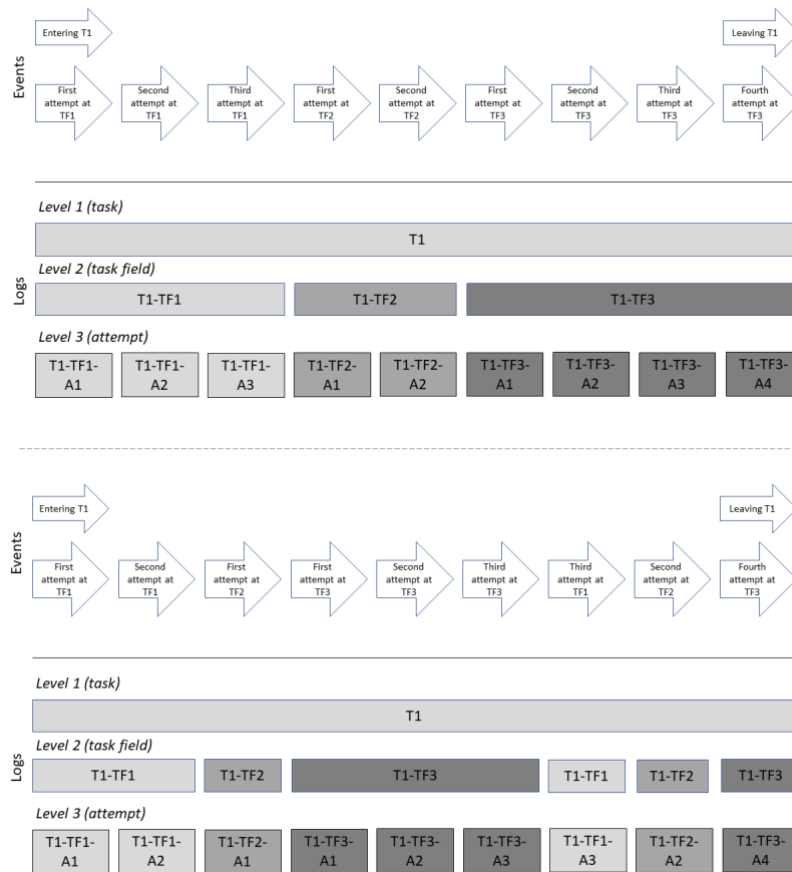


Figure 5 presents two cases of interaction within our definition of timeframes. In the first case (upper panel), the learner entered Task 1 (T1), at which point was also the start of the time period at all three levels (task, task field, and attempt). This learner then moved onto working on Task Field 1 (TF1) which they attempted three times. The first check attempt (A1) would mark the end of the first attempt period for this task field as well as the start of the next attempt (i.e., T1-TF1-A2). The same applied to the third submission (T1-TF1-A3). After that, which was also the end of TF1, they worked on TF2 (the second task field). We applied the same principle in defining the time windows to subsequent task fields and attempts. Eventually, when the learner left the task, that time point also marked the end of the time windows at all three levels. We



acknowledge that this analysis assumes that the learner stayed on task, that is, did not take a break. Note, though, that the definition of time windows is able to handle cases where the learner did not attempt items in a sequential order and returned to previously attempted items. In theory, this could happen at the task level as well, but we did not observe any such moves. We present another example case that was more complex in that the learner moved back and forth between different task fields within a task (lower panel of [Figure 5](#)).

Given the definition of time frames, we were able to generate learning process variables from the system logs. We wanted to capture four different, yet related, types of information in our data. To identify relevant variables, we drew on Chen et al.'s (2018) study as reviewed above. We also included additional variables to shed light specifically on language learning. We included some similar variables but at different levels (e.g., Total Time on Task *vs.* Time on Task Field) to examine potential differences in accounting for students' interaction with the system. Importantly, having data extracted from more than one level could further minimize measurement errors as a result of the statistical technique that we employed (see Data Analysis below). [Table 1](#) summarizes all variables with their definitions.

**Table 1***List of Learning Process Variables*

<b>Variable</b>	<b>Definition</b>
Level 1 (Task)	
1. Total time on task	Total time (in seconds) spent on all tasks
2. Finish time relative to first in class	Average finish time difference (in days) relative to the first submission in the class
3. Finish rank in class	Average finish rank for submissions relative to classmates
4. Tasks attempted	Proportion of tasks in this chapter attempted
5. Fill-in-the-blank task attempted	Proportion of fill-in-the-blank tasks attempted
6. Short-answer task attempted	Proportion of short-answer tasks attempted
7. True-or-false task attempted	Proportion of true-or-false tasks attempted
8. Task interval	Average time between all tasks
Level 2 (Task field)	
9. Task fields attempted	Number of task fields attempted in all tasks (not left empty)
10. Task fields not attempted	Proportion of task fields not attempted (left empty)
11. Task field interval (overall)	Average time for all task fields
12. Correct attempts per task field	Proportion of correct submissions for each attempted task field over all attempted task fields
13. Time per task field	Average time spent per task field
Level 3 (Attempt)	
14. Correct attempt	Proportion of correct attempts for each task field
15. Incorrect attempt	Proportion of incorrect attempts for each task field
16. Total attempts	Total number of attempts for each task field
17. First correct	Proportion of correct answer on first attempt
18. Attempt intervals (First to second)	Average time (in seconds) between the first and second attempt for all task fields
19. Proportion of specific feedback	Proportion of specific ( <i>vs.</i> general correct/incorrect) feedback over all attempts

## Design and Procedure

At the beginning of the study, teachers were introduced to the system's functionality. The students were randomly assigned to one of two condition groups, with the number of students in each condition group balanced in every school class. For a given chapter, half of the students were randomly assigned to the intervention group, and half of the students were assigned to the control group. Again, those students who were assigned to the intervention group received interactive scaffolding feedback, whereas the students in the control group received general feedback only on the correctness of their answers. Before and after every chapter, the students completed an independent pre- and post-test that measured their grammatical knowledge.

## Data Analysis

### *Data processing*

In the spirit of open science, we have made our raw data and data analysis code available on Open Science Framework (OSF) at <https://osf.io/39v4z/>. Before addressing the research questions, we inspected the descriptive statistics of the data for normality and outliers. Since the data were collected from naturalistic contexts, caution was exercised in order to maximize reliability of the data. A visual inspection of plots suggested issues with outliers in a number of variables. To address these issues, we treated outliers in the data by means of winsorization (Wilcox, 2005). This procedure replaced extreme values with less extreme values. The trim quantile was specified as .10, meaning that observations below and above than the fifth and 95th percentiles were replaced by values at the fifth and 95th percentiles respectively. We then centered and scaled (standardized) all process variables so that they were on the same scale for comparability. Following Meurers et al. (2019), for the pre-/post-test scores, we computed a gain score for each participant representing their improvement in grammatical knowledge from pre- to post-test.

### **RQ1**

To address the first research question, concerning the relationship between learning process and products, we used a regression analysis. Despite the nested data structure (students nested within classes), the intra-class correlation (ICC) at .08 was below the typical range from .10 to .25 in educational contexts (Hedges & Hedberg, 2007). Therefore, we modeled the data with a single-level regression analysis to avoid convergence issues, striking a balance between more accurate estimates and practicality in modeling.

The outcome was specified as the gain score on the pre- and post-tests. For the predictors, we reduced the dimensions of learning process variables by performing an exploratory factor analysis (EFA). We engaged in two rounds of EFA using the `fa()` function in R's `psych` package (version 1.8.4). In determining the number of factors to be extracted, we used the `eigen()` function in the `nFactors` package (version 2.3.3) to compute eigen values for the factors to be extracted, representing the amount of variance explained by them. We plotted these eigen values onto a scree plot to identify the point of inflexion indicating the point where extracting more factors would no longer proportionally increase the amount of variance explained. This point of inflexion represents the appropriate number of factors for a given data set (e.g., Cattell, 1966). We used maximum likelihood<sup>1</sup> as the factoring method, which allows generalization of the results (e.g., Loewen & Gonulal, 2015). To ensure interpretability of the results, we used an oblimin rotation to rotate the resulting matrices. An oblimin rotation is a kind of oblique rotation which allows factors to correlate, and hence is preferred (e.g., Brown, 2015). During the end of the first round, we inspected the factor loadings for each variable across the factors, representing the strength of association between the variable and the factor in question. We discarded variables that did not load very well to any of the factors extracted. We used the threshold of .63 because only loadings at and above this level are considered very good (Comrey & Lee, 1992), rendering a purer measure of the theoretical construct we were interested in. We then engaged in the final, second round of factor analysis by repeating the same procedure. We then estimated factor scores from the EFA for each participant using regression. By doing this, we used these latent measures of the learning process variables as the predictors of the main multiple regression analysis as predictors.

In addition to the factors from the EFA, we also included the variables that were discarded after the first round of EFA. These variables could be important independently in accounting for eventual learning even though they did not load onto any of the factors. Moreover, we included pre-test score and condition group (intervention (1) vs. control (0), dummy coded) as covariates. In the regression analysis, we engaged in a backward model selection process. At each step of the model selection process, we removed the predictor that had the highest  $p$ -value until all predictors were significant or when the adjusted  $R^2$  value started to decrease. The final model thus represented the model accounting for the most variance in gain score while adjusted for model complexity.

### **RQ2**

To address the second research question, concerning clusters of learners based on significant process variables, we engaged in a cluster analysis. We examined the extent to which we were able to identify groups of students who share similar learning process patterns based on the significant features identified in the regression analysis. We used the k-means approach because we did not assume any hierarchical nature of the data (Staples & Biber, 2015). We used the `kmeans()` function in R's stats package (version 3.6.0) with the number of random sets (`nstart`) specified as 40. In determining the number of clusters, we compared the proportion of between-cluster variance. The goal was to identify a point where an increase in number of clusters would not substantially improve the proportion of between-clusters variance explained. We tested solutions of two to six clusters. We report and interpret the optimal solution.

### **RQ3**

With regard to the third research question about the relationship between provision of specific feedback and learning process patterns, we performed a chi-square test of independence to examine the potential relationship between learning conditions, which is also our experimental manipulation (i.e., condition group: intervention vs. control), and learning process patterns (learner cluster). A chi-square test reveals the extent to which two variables, condition group and process cluster in the present case, are independent from each other. In other words, could learners in the intervention group, for example, tend to demonstrate certain process patterns (or fall into a certain cluster)?

## **Results**

### **RQ1: To What Extent Can Learning Process Variables, as Extracted from System Logs, Directly Account for Learning Outcomes?**

In reducing the dimensions of the process variables, the final EFA model was a three-factor solution from 14 observed variables. In other words, five process variables were removed from the final EFA model. In [Table 2](#), we present the factor loadings of each of the process variables, communalities ( $h^2$ ) (indicating the amount of variance explained for a variable across all factors), and (cumulative) proportions of variance explained by each factor. The factor loadings ranged from .78 to 1.01, demonstrating the factors were a good representation of the indicators (i.e., the process variables). The cumulative proportion of variance was high at .85, representing that the variance in the data was adequately explained. Finally, the communality values ranged from .56 to 1.00, meaning that information carried by the indicators was well explained by the factors. In sum, we considered that the current analysis adequate in capturing the patterns underlying the process variables. In order to label the factors in a meaningful manner for interpretation purposes, we inspected the constituents of each factor. For Factor 1, all constituents were related to the number of attempts at different temporal levels. Therefore, we labeled Factor 1 as *effort*, representing the learner's level of engagement with the system. For Factor 2, since all components were based on the accuracy of the learner responses, we labeled it *accuracy-focus*, indicating how accurate a learner was when completing the tasks within the system. Finally, we labeled Factor 3 *time on task* because the three indicators revealed how much time was spent in different time frames.

**Table 2***Rotated Loadings for the Learning Process Variables*

	<b>Effort</b>	<b>Accuracy-Focus</b>	<b>Time on Task</b>	<b>h<sup>2</sup></b>
Total attempts	1.00			0.98
Tasks attempted	0.97			0.96
Task field attempted	0.93			0.92
Fill-in-the-blanks tasks attempted	0.92			0.82
Short-answer tasks attempted	0.81			0.76
True-or-false tasks attempted	0.78			0.59
Correct attempts per task field		1.00		1.00
Incorrect attempts		-0.95		0.94
Correct attempts		0.93		0.84
Task fields not attempted		-0.82		0.70
Proportion of first correct		0.73		0.56
Time per task field			1.01	1.00
Task interval			1.00	0.98
Total time on task			0.79	0.91
Proportion of variance explained		0.37	0.29	0.20
Cumulative proportion of variance explained		0.37	0.66	0.85

In terms of the regression analysis, the final model had an adjusted  $R^2$  value of 0.37, suggesting that 37% of the variance in gain score can be attributed to the predictors in the model. This amount represents a medium effect size in L2 research (Plonsky & Ghanbar, 2018). The model indicated that both the factor of accuracy-focus and the process variable of finish time were significant predictors. In general, the more accurate a learner was in the system, the larger gains they demonstrated from pre- to post-tests. In terms of finish time, early submissions were positively associated with gains. The model summary is presented in [Table 3](#).

**Table 3***Regression Model Summary*

	<b>Estimates [95% CI]</b>	<b>p</b>
Intercept	7.05 [6.25, 7.86]	<.001
Pre-test score	-3.03 [-3.62, -2.44]	<.001
Condition group - Intervention	2.05 [0.91, 3.20]	<.001
Accuracy	0.63 [0.05, 1.21]	.03
Finish time	-0.89 [-1.67, -0.11]	.03
Multiple R <sup>2</sup> / Adjusted R <sup>2</sup>	.38 / .37	
ANOVA	F (4, 200) = 30.56, $p < .001$	

**RQ2: Can We Meaningfully Distinguish Clusters of Learners Based on the Learning Process Variables?**

In addressing RQ2, we report the four-cluster solution based on the proportion of between-clusters variance explained (see Table 4). In Table 5, we present the descriptive statistics of the four clusters of learners in regard to the variables significant in the regression analysis. We also visualize these clusters in Figure 6.

**Table 4***Between-clusters Variance Explained by Different Solutions*

	Between-clusters variance explained
Two-cluster solution	46.9%
Three-cluster solution	65.1%
Four-cluster solution	75.7%
Five-cluster solution	82.6%
Six-cluster solution	85.3%

**Table 5***Descriptive Statistics of the Significant Process Variables by Clusters*

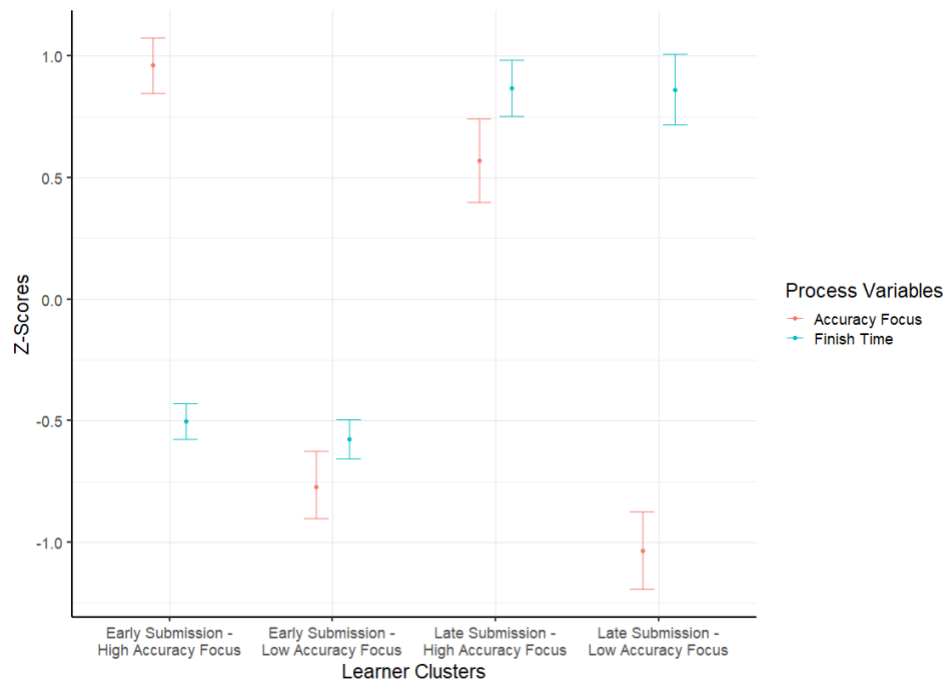
Cluster	Process Variable / Factor	
	Finish time	Accuracy-Focus
	Mean (SD)	
	[95% CIs]	
1 (late submission & low accuracy) (n = 32)	0.86 [0.72, 1.01]	(0.40) -1.04 [-1.20, -0.87] (0.46)
2 (early submission & high accuracy) (n = 69)	-0.50 [-0.58, -0.42]	(0.33) 0.96 [0.86, 1.07] (0.46)
3 (late submission & high accuracy) (n = 35)	0.87 [0.74, 1.00]	(0.37) 0.57 [0.39, 0.75] (0.53)
4 (early submission & low accuracy) (n = 69)	-0.58 [-0.66, -0.49]	(0.34) -0.77 [-0.91, -0.64] (0.57)

*Note:* Finish time and accuracy-focus were standardized (mean of 0 and z-transformed).



Figure 6

## Clustering by Process Variables



A number of observations are noteworthy. First, the cluster analysis returned a  $2 \times 2$  structure in terms of the variables this analysis was based on (accuracy-focus and finish time). Learners in clusters 1 and 3 submitted their work relatively late (0.87 and 0.84) but differed in their accuracy (-1.04 for Cluster 1 and 0.57 for Cluster 3). Clusters 2 and 4 submitted early relative to their peers (-0.50 and -0.58 for finish time). The difference between these two clusters is the accuracy-focus where Cluster 2 had a high accuracy-focus (0.96), and Cluster 4 had low accuracy-focus (-0.77). In sum, these clusters could be described in terms of early *vs.* late submission and high *vs.* low accuracy-focus.

### RQ3: To What Extent Does Specific Feedback Relate to the Learning Process Cluster Characteristics of a Learner?

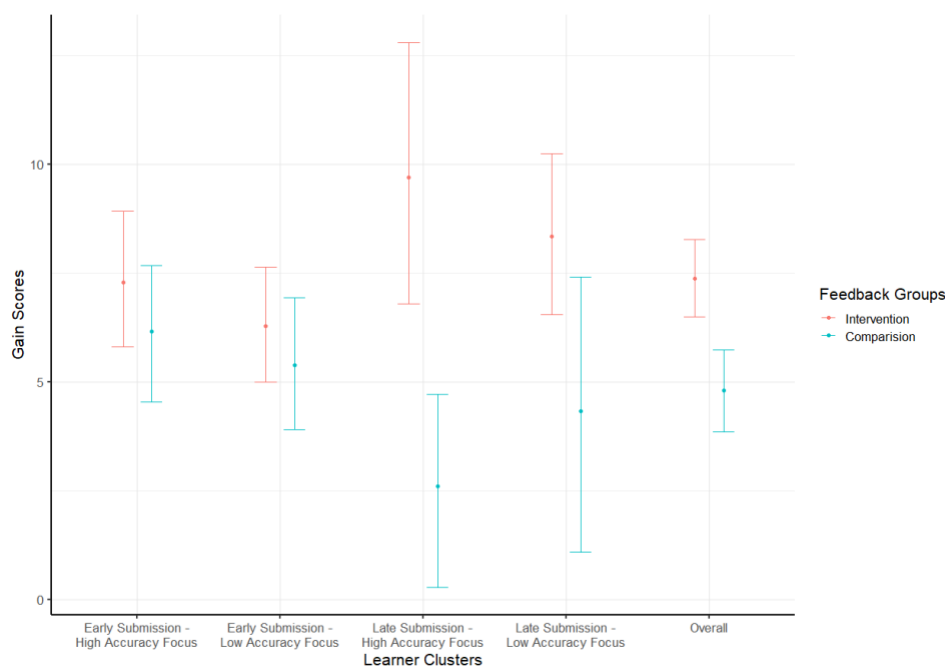
Addressing RQ3, regarding the relationship between learning conditions (provision of specific feedback) and process patterns, we moved from between-cluster comparisons to within-cluster observations. We inspected the differences between learners in the intervention and the control group within a given cluster. Despite the fact that learners demonstrated similar process patterns (i.e., within the same cluster), we observed that the number of learners belonging to the intervention group and to the control group was telling. Although assignment of condition groups was random ( $n = 104$  in the intervention,  $n = 101$  in the control), the proportion of learners from each group was not always equal or similar within a given cluster (see Table 6). A chi-square test of independence confirmed that condition group and cluster were not independent ( $\chi^2(3) = 9.23, p = .03$ ), indicating tendency for learners in the intervention group to demonstrate particular process patterns as captured by the clustering. For example, Cluster 1 (late submission – low accuracy-focus) had more learners belonging to the intervention group than to the control group ( $n = 20$  *vs.* 12). In contrast, Cluster 3 (late submission – high accuracy-focus) had more than twice as many learners in the control group than the intervention group ( $n = 25$  *vs.* 10).

**Table 6***Descriptive Statistics of Gain Scores by Clusters and Condition Groups*

Cluster		Learning Products: Pre- /post-test gain score		Group comparison on gain score
		Mean (SD)		
		[95% CIs]		
1 (late submission & low accuracy)	Intervention (n = 20)	8.35 [6.21, 10.50]	(4.58)	$t(19) = 2.03, p = .06,$ $d = 0.79$ 95% CIs [0.04, 1.52]
	Control (n = 12)	4.33 [0.60, 8.07]	(5.88)	
2 (early submission & high accuracy)	Intervention (n = 38)	7.29 [5.59, 8.99]	(5.18)	$t(66) = 0.96, p = .34,$ $d = 0.23$ 95% CIs [-0.25, 0.70]
	Control (n = 31)	6.16 [4.49, 7.84]	(4.57)	
3 (late submission & high accuracy)	Intervention (n = 10)	9.70 [5.81, 13.60]	(5.44)	$t(18) = 3.38, p = .003,$ $d = 1.21$ 95% CIs [0.41, 1.99]
	Control (n = 25)	2.60 [0.12, 5.09]	(6.02)	
4 (early submission & low accuracy)	Intervention (n = 36)	6.28 [4.85, 7.71]	(4.23)	$t(65) = 0.83, p = .41,$ $d = 0.20$ 95% CIs [-0.27, 0.67]
	Control (n = 33)	5.39 [3.76, 7.03]	(4.61)	

**Figure 7**

Visualization of Gain Scores by Learner Clusters and Feedback Condition Group



Given the relationship between condition group and process patterns, we engaged in exploratory analysis to further inspect the relationship between clusters, gain scores, and condition group. While the emergent nature of the clusters brings with it the fact that we did not have particular hypotheses about them, we found that the explorative analysis and descriptive statistics provide potentially valuable insights for understanding the data. Specifically, while learners in the same cluster showed similar process patterns and baseline (pretest scores), for some clusters, we observe markedly different learning outcomes for the different condition groups (see Table 5). For example, in Cluster 3 (late submission – high accuracy-focus), the intervention group gained an average of 9.70 points from pre- to post-test, compared with learners in the control group, who showed only a mean gain score of 2.60. Similarly, in Cluster 1 (late submission – low accuracy-focus), the intervention group also appeared to have outperformed the control group. Although the simple effect of condition group in the regression analysis as well as our previous publication revealed the fact that the intervention group generally obtained larger gain scores than the control group, the key finding here pointed to a moderating effect of intervention group on gain score that was based on learner cluster. In other words, receiving specific feedback was generally useful, but more so or only when the learner belonged to a particular cluster (i.e., demonstrating certain process patterns). As such, we performed a follow-up ANOVA where the dependent variable was the gain score, and the independent variables were condition group, cluster and their interaction term. Results confirmed a significant interaction between condition group and cluster ( $F(3) = 3.32, p = .02$ , partial  $\omega^2 = .03$ , 90% CIs [0.00, 0.07]). We followed up this interaction by performing pair-wise  $t$ -tests for each cluster (see Table 6). Since the sample size was relatively small in the pairwise comparisons, we relied on the effect sizes (Cohen's  $d$  and its 95% CIs) for our interpretation. In particular, meaningful differences between the intervention and control groups were only found in clusters 1 (late submission & low accuracy) and 3 (late submission & high accuracy). We visualize these differences in Figure 7. This moderating effect points to the need to consider both learning conditions (provision of specific feedback) and learning processes (learner clusters based on process variable) in understanding learners' success in eventual learning.

## Discussion

In this study, we analyzed system logs extracted from an ICALL system as a proxy of language learning processes. We also related this wealth of information to learning conditions (i.e., provision of specific grammar feedback) and eventual learning outcomes. Early submissions and a high focus on accuracy when working with the system predicted higher gain scores on the pre-/post-tests. Four learner types (clusters) were identified based on these significant variables. We further found that learning conditions interacted with these clusters in accounting for eventual learning outcomes. In particular, specific feedback was able to help in most cases (Meurers et al., 2019), but only for learners who demonstrate specific learning process patterns.

First and foremost, we have shown that system logs can be used to generate meaningful learning analytics which in turn offers insights into second language learning as a process (e.g., Godfroid, 2019; Leow, 2015). It is a promising route for SLA and CALL research. In principle, all systems that incorporate some interactive features can generate system logs. Although our results remain silent on the cognitive aspects of language processing, these logs and data represented how learners used the system and performed in the tasks. Through exploring and understanding these data, researchers and system designers are in a better position to create conditions that could be more conducive to learning (Mangaroska & Giannakos, 2018; Siemens & Baker, 2012). For example, blind guessing can be manifested in the large number of attempts in relatively short periods of time. When these learners also did not achieve high learning gains, one account would be that they did not engage in deep processing of the feedback provided for them. A specific example of blind guessing we observed in our logs concerns a student entering the answer *say* in a blank-filling task where the base form (*say*) was given in parentheses. The system gave the feedback “An expression like 'yesterday' tells you that something happened in the past and is over now. It is a signal word for the simple past.”. Two seconds later, the learner submitted the very same answer again. Then, after 14 seconds, this learner submitted the correct answer *said*. Potentially, scrutinizing information such as this example can help researchers understand SLA as a process.

Although we did not systematically zoom in to this level of detail, our macro level of analysis also offers fruitful insights into the nature of the feedback advantage we observed in our previous publication (Meurers et al., 2019). Had we ignored learning processes altogether, we would not have found the interaction between learner cluster and provision of specific corrective feedback. The results that students submitting late benefited most from specific feedback is intriguing, though our approach remains silent on the exact cause of this interaction—an interesting avenue to pursue in future research.

Further, the importance of accuracy-focus in the system found in this study, if replicated consistently, underscores the need for researchers and system designers to explore effective strategies to promote more deep processing (e.g., Hattie & Timperley, 2007). For example, could some gamification features be incorporated to promote deliberation and focus on accuracy (e.g., point rewards considering both correctness and number of attempts)? Additionally, research should focus on the reasons learners submit late relative to their peers and consequently how they benefit from the specific feedback in this situation. Again, studying system logs is then an important means to lay the foundation for further research into optimizing language learning experiences in a CALL environment.

In terms of the important process variables in accounting for learning, we confirmed and extended the findings reported in Chen et al. (2018). These authors identified the importance of finish time and time on task in accounting of ultimate achievement. With our dataset, we also identified the role of finish time, but time on task was not a significant predictor in our analysis. The lack of association between achievement and time on task is a finding compatible with previous SLA research that found more practice did not necessarily entail better learning (e.g., Cerezo, 2016, Cerezo et al., 2016). Rather, accuracy focus explained unique variance in our data. Potential accounts for the difference between our results and those reported in Chen et al. (2018) include the different learning domains investigated (physics vs. EFL) and levels of learners (university vs. high school learners). Importantly, we extended their work by showing that

individual variables need to be considered in tandem with not only other process variables, but also learning conditions, which in our case was the provision of specific grammar feedback. The relationships between learning condition, process, and outcomes deserve much more work.

In terms of our analytical approach, we took a step towards understanding how system log data can be extracted and analyzed meaningfully. Such distilled data could be useful for stakeholders such as teachers, researchers, students, and system designers. We showed how one can obtain information from system logs to process variables and linked them to eventual learning outcomes, essentially turning such data to meaningful learning analytics. The analysis revealed that there can be different paths to success, and that it is not sufficient to only consider information at one level (e.g., task) or one dimension (e.g., submission time). On the contrary, the factor scores resulted from the factor analysis captured both the breadth and depth of our data. In terms of breadth, these three factors compare favorably to traditional, summative indices that often only focus on performance (e.g., percentage of correct answers by item or by student). In terms of depth, these three factors crystalized rich information from a large number of fine-grained measures. Given the different learning paths, we emphasize the importance of looking at different dimensions of students' use of the system.

We acknowledge that our quantitative data relied on subjective interpretations. Although we drew on our years of teaching experience in interpreting our data, we were unable to confirm the exact behavior, learning strategies and/or processes in the course of students' interaction with the system. Therefore, the present study remains silent on *why* there exists different learner clusters beyond showing that they can be associated with the provision of specific feedback. A broader discussion needs to ensure that detailed and qualitative insights become available through, for example, interviews and learner diary entries. This line of research also needs to be pushed further by making more explicit records of interpretable learning process markers available, such as uptake to specific feedback. In terms of the granularity of the analyses, we considered cases when learners submitted an answer to the feedback mechanism. In future work, for example, the analyses could be refined by implementing keystroke logging to even better understand the edit patterns of learners. Furthermore, we did not look at the role of teacher feedback, but instead focused on learner interaction with automatic feedback. The diagnostic information currently used in an offline analysis can also be made usable in the system in the form of a teacher dashboard, with a student-at-risk detection module warning teachers when the system detects potential, struggling learners.

## Conclusion

We explored the connection between learning conditions, processes, and outcomes by analyzing and interpreting detailed learner interaction logs. The logs were collected in the context of a randomized controlled field trial using the FeedBook system implemented in authentic and ecologically valid classrooms. We showed that accuracy-focus, a factor obtained in exploratory factor analysis, and finish time were two significant predictors of grammar learning outcomes. In a cluster analysis, we showed that different learning paths can lead to similar learning outcomes and that provision of specific feedback interacted with learner cluster when linking to ultimate achievement.

Our focus on language learning as a process highlights the wealth of information this line of research could offer. We observed that with specific grammar feedback, many learners are eventually able to submit a correct answer; but what differs is the path towards reaching the solution. Without integrating learning analytics, one in essence pursues a blackbox approach, in which one does not know which learning path the learners took and which issues they had reaching their solutions, thus lacking crucial diagnostic information needed to understand the connection between learning conditions, processes, and outcomes. The present paper represents a step towards understanding by interpreting system logs as a window onto language learning processes. Future research arguably needs to employ various different methodologies to identify and interpret a broad range of information on what happens during the learning process. Advancing our understanding of the processes and the results they lead to then can also pave the way for better interventions and system design.

## Acknowledgments

The authors are grateful to the FeedBook project team, the LEAD Research Network at the University of Tübingen, and the Second Language Studies PhD program at Michigan State University. For constructive feedback on this paper, we would like to thank Ramon Ziai, Glenn Davis, and Caitlin Cornell, the audience members at the Second Language Research Forum (SLRF) 2019 at East Lansing, Michigan and the reviewers.

## Note

<sup>1</sup> Since maximum likelihood has multivariate normality assumptions and some of our behavioral variables were not normally distributed, we refitted our final solution using the sandwich standard error estimates as implemented in the EFAutilities package for the non-normal distribution (Zhang et al., 2012). This model arrived at the same conclusion, also echoing the observation by Loewen and Gonulal (2015, p. 193) that “the practical differences between ... extraction methods are frequently negligible.”

## References

- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245–276. [https://doi.org/10.1207/s15327906mbr0102\\_10](https://doi.org/10.1207/s15327906mbr0102_10)
- Cerezo, L. (2016). Type and amount of input-based practice in CALI: The revelations of a triangulated research design. *Language Learning & Technology, 20*(1), 100–123.
- Cerezo, L., Caras, A., & Leow, R. P. (2016). The effectiveness of guided induction versus deductive instruction on the development of complex Spanish gustar structures: An analysis of learning outcomes and processes. *Studies in Second Language Acquisition, 38*(2), 265–291. <https://doi.org/10.1017/S0272263116000139>
- Chen, X., Breslow, L., & DeBoer, J. (2018). Analyzing productive learning behaviors for students using immediate corrective feedback in a blended learning environment. *Computers and Education, 117*, 59–74. <https://doi.org/10.1016/j.compedu.2017.09.013>
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Erlbaum.
- Elgort, I., Brysbaert, M., Stevens, M., & Van Assche, E. (2018). Contextual word learning during reading in a second language: An eye-movement study. *Studies in Second Language Acquisition, 40*(2), 341–366. <https://doi.org/10.1017/S0272263117000109>
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let’s not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71. <https://doi.org/10.1007/s11528-014-0822-x>
- Godfroid, A. (2019). Investigating instructed second language acquisition using L2 learners’ eye-tracking data. In R. P. Leow (Ed.), *The Routledge handbook of second language research in classroom learning* (pp. 44–57). Routledge. <https://doi.org/10.4324/9781315165080-4>
- Godfroid, A. (2020). Sensitive measures of vocabulary knowledge and processing. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (1st ed., pp. 433–453). Routledge. <https://doi.org/10.4324/9780429291586-28>
- Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoon, H.-J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition, 21*(3), 563–584. <https://doi.org/10.1017/S1366728917000219>
- Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in

- incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, 35(3), 483–517. <https://doi.org/10.1017/S0272263113000119>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, 29(1), 60–87. <https://doi.org/10.3102/0162373707299706>
- Heift, T. (2010). Developing an intelligent language tutor. *CALICO journal*, 27(3), 443–459. <https://doi.org/10.11139/cj.27.3.443-459>
- Heift, T. (2019). A longitudinal observation of technology-mediated feedback for L2 learners of German. *ITL-International Journal of Applied Linguistics*, 170(2), 154–179. <https://doi.org/10.1075/itl.19009.hei>
- Hui, B. (2020). Processing variability in intentional and incidental word learning: An extension of Solovyeva and DeKeyser (2018). *Studies in Second Language Acquisition*, 42(2), 327–357. <https://doi.org/10.1017/S0272263119000603>
- Leow, R. (2015). *Explicit learning in the L2 classroom: A student-centered approach*. Routledge. <https://doi.org/10.4324/9781315887074>
- Loewen, S., & Gonulal, T. (2015). Exploratory factor analysis and principal components analysis. In L. Plonsky (Ed.) *Advancing quantitative methods in second language research* (pp. 182–212). Routledge.
- Mangaroska, K., & Giannakos, M. N. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*, 12(4), 516–534. <https://doi.org/10.1109/TLT.2018.2868673>
- Marsden, E., Mackey A., & Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS Repository of Instruments for Research into Second Languages* (pp. 1–21). Routledge. <https://doi.org/10.4324/9780203489666>
- 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, 161–188. <https://doi.org/10.1017/S0267190519000126>
- Plonsky, L., & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal*, 102(4), 713–731. <https://doi.org/10.1111/modl.12509>
- Robinson, P., Mackey, A., Gass, S. M., & Schmidt, R. (2013). Attention and awareness in second language acquisition. In *The Routledge handbook of second language acquisition* (pp. 265–285). Routledge. <https://doi.org/10.4324/9780203808184-24>
- Rudzewitz, B., Ziai, R., De Kuthy, K., & Meurers, D. (2017). Developing a web-based workbook for English supporting the interaction of students and teachers. In E. Volodina, G. Grigonytė, I. Pilán, K. N. Björkenstam, & L. Borin (Eds.), *Proceedings of the joint workshop on NLP for computer assisted language learning and NLP for language acquisition* (pp. 36–46). LiU Electronic Press.
- Rudzewitz, B., Ziai, R., De Kuthy, K., Möller, V., Nuxoll, F., & Meurers, D. (2018). Generating feedback for English foreign language exercises. In J. Tetreault, J. Burstein, E. Kochmar, C. Leacock, & H. Yannakoudakis (Eds.), *Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications* (pp. 127–136). Association for Computational Linguistics.



<https://doi.org/10.18653/v1/W18-0513>

- Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252–254). ACM. <https://doi.org/10.1145/2330601.2330661>
- Staples, S., & Biber, D. (2015). Cluster analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 243–274). Routledge. <https://doi.org/10.4324/9781315870908-11>
- Wilcox, R (2005). *Introduction to robust estimation and hypothesis testing*. Elsevier.
- Zhang, G., Preacher, K. J., & Jennrich, R. I. (2012). The infinitesimal jackknife with exploratory factor analysis. *Psychometrika*, 77, 634–648. <https://doi.org/10.1007/s11336-012-9281-5>
- Ziai, R., Rudzewitz, B., De Kuthy, K., Nuxoll, F., & Meurers, D. (2018). Feedback strategies for form and meaning in a real-life language tutoring system. In I. Pilán, E. Volodina, D. Alfter, & L. Borin (Eds.), *Proceedings of the 7th workshop on NLP for computer assisted language learning* (pp. 91–98). LiU Electronic Press.

## About the Authors

Bronson Hui is Assistant Professor of Second Language Acquisition at University of Maryland, College Park. His research interests include second language acquisition, applied psycholinguistics, and quantitative methods. His work has appeared or will appear in *Studies in Second Language Acquisition*, *Language Learning*, *Modern Language Journal*, and *Applied Psycholinguistics*, among other venues.

**E-mail:** [bhui@umd.edu](mailto:bhui@umd.edu)

Björn Rudzewitz is a postdoctoral researcher at the University of Tübingen, Germany, specializing in Learning Analytics in the context of Intelligent Computer-Assisted Language Learning and one of two main developers of the FeedBook system presented here.

**E-mail:** [br@sfs.uni-tuebingen.de](mailto:br@sfs.uni-tuebingen.de)

Detmar Meurers is Professor of Computational Linguistics at the University of Tübingen, Germany, a steering board member of the LEAD Research Network in Empirical Educational Science, and a member of the Machine Learning Cluster of Excellence, EXC number 2064/1 – Project.

**E-mail:** [dm@sfs.uni-tuebingen.de](mailto:dm@sfs.uni-tuebingen.de)