Project - Datasets

Machine Learning for Behavioral Data (CS-421)

February 17, 2026

Start-up Presentations

Two EdTechCollider start-ups have provided us with real-world data:

- GoGymi
- Lernnavi

GoGymi

- Platform with an AI tutor to prepare for the gymnasium exam

GoGymi In Numbers

50+	78+	820+	2,000+
partner schools	teachers	learners	exercises

01

Personal AI Tutor

Your digital learning companion is available 24/7 - for questions, explanations and feedback. You can chat and talk to the AI tutor and receive tailor-made answers.

02

AI-Based Essay Correction

Upload your handwritten texts and immediately receive detailed feedback for improvement and correction according to the official correction scheme.

03

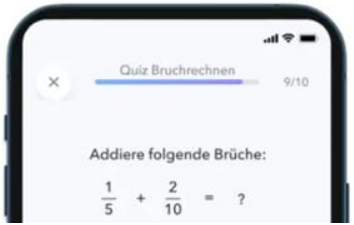
Exercises & Solutions

With online exercises, learners can deepen their knowledge and gain access to step-by-step solutions to all old exams from the last 10 years.

04

Learning Analytics

Our intelligent system recognizes your strengths and weaknesses and automatically adjusts your study plan. This gives you detailed analyses of your learning activity and assessments of your exam readiness.



GoGymi
Data Overview

- Scores obtained by students in practice exams
- Graded essays submitted by students
- User interaction with the intelligent AI-based chatbot
- Activity logs of users

Name	Description
students.csv	A set of students registered in GoGymi.
teachers.csv	A set of teachers registered in GoGymi.
all_scores.csv	The score (in percentages) obtained by each student in each exam of each course, as well as timestamp.
text_results.csv	The points obtained by each student in the text exam questions of the text comprehension course, as well as the maximum grade per question and timestamp. With 1 question per row.
math_results.csv	The points obtained by each student in the math questions of each exam of the math course, as well as the maximum grade per question, timestamp, and if they requested a hint, where a hint was available.
essay_results.csv	The scores of the submitted essays obtained by each student for each course and text type. Contains the text submitted by the user, as well as a set of measured metrics.
activity.csv	The list of logs of user interactions with the tool, along with timestamps.
gymitrainer.csv	The list of user interactions with the intelligent AI-based chatbot.
math_questions.csv	The list of math questions in each exam.
text_questions.csv	The list of textual questions in each exam.
texts.csv	The list extracts provided for each exam which the students answer questions on. The extracts are the basis of the text questions for students. There is exactly one per each exam.

GoGymi | Project Ideas

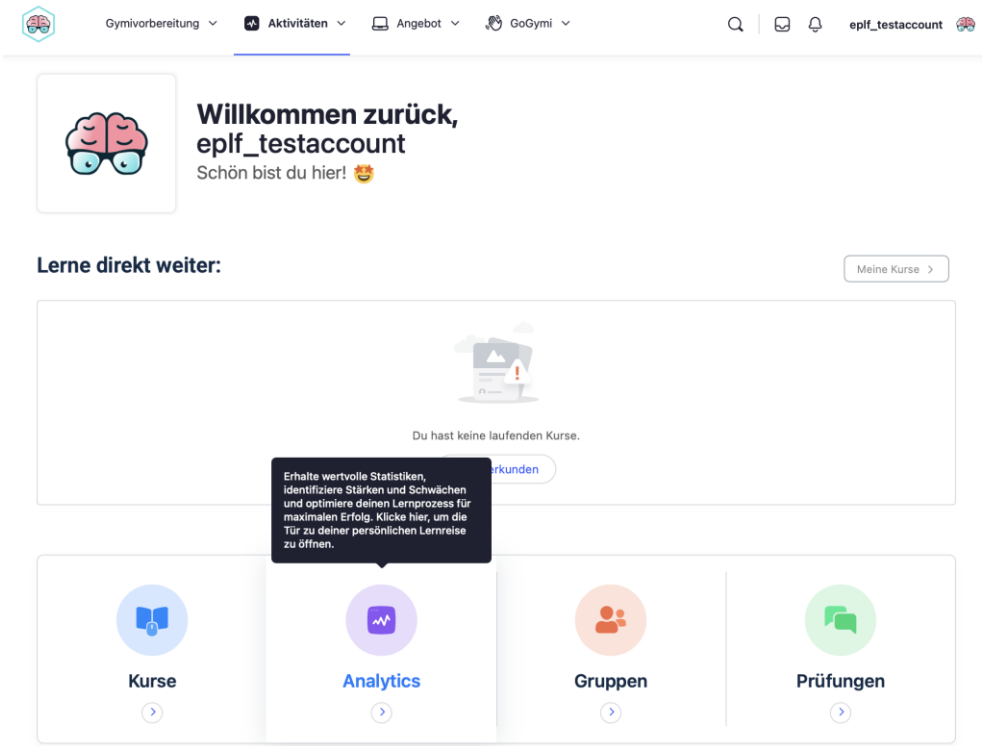
- **Time series analysis** of students.
- Does usage of the **intelligent chatbot** lead to better learning outcomes?
- Which types of **activities** are correlated with higher obtained results?
- Which behavioral groups / clusters of users show the most or least **improvement** over time?
- Can the learning results be predicted based on **automatic analysis** of the submitted **essays**?

GoGymi | Demo

https://gogymi.ch/ggy/

Username: eplf_testaccount

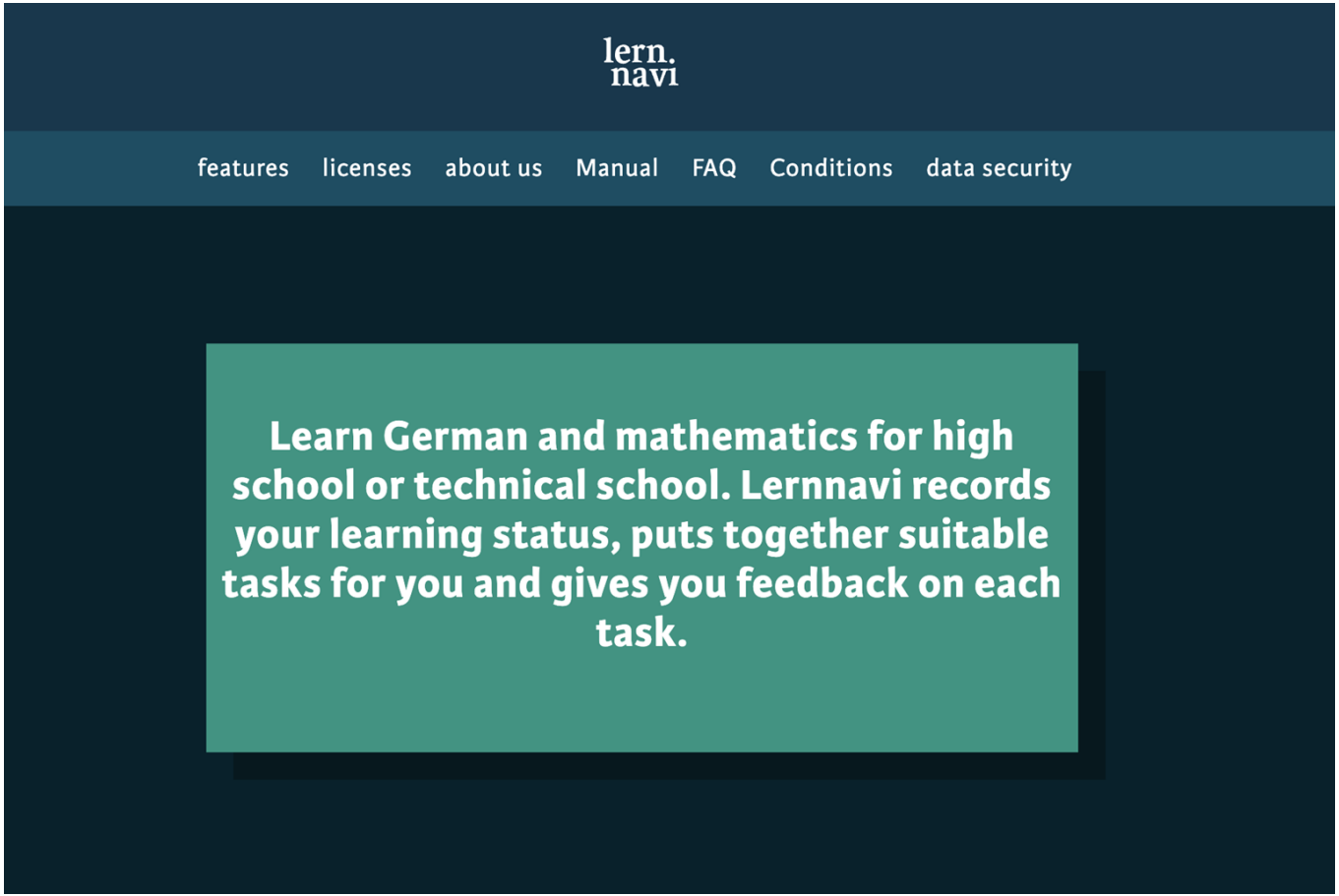
Password: ml4ed2026



Lernnavi

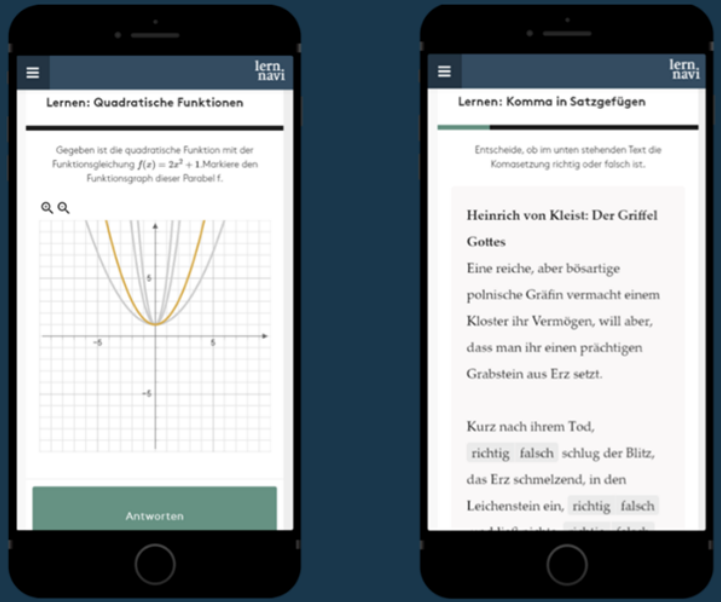
- Platform (ITS) to learn Math and German.
- Targeted to high school students.

- More than 31,000 students and 11,000,000 events.

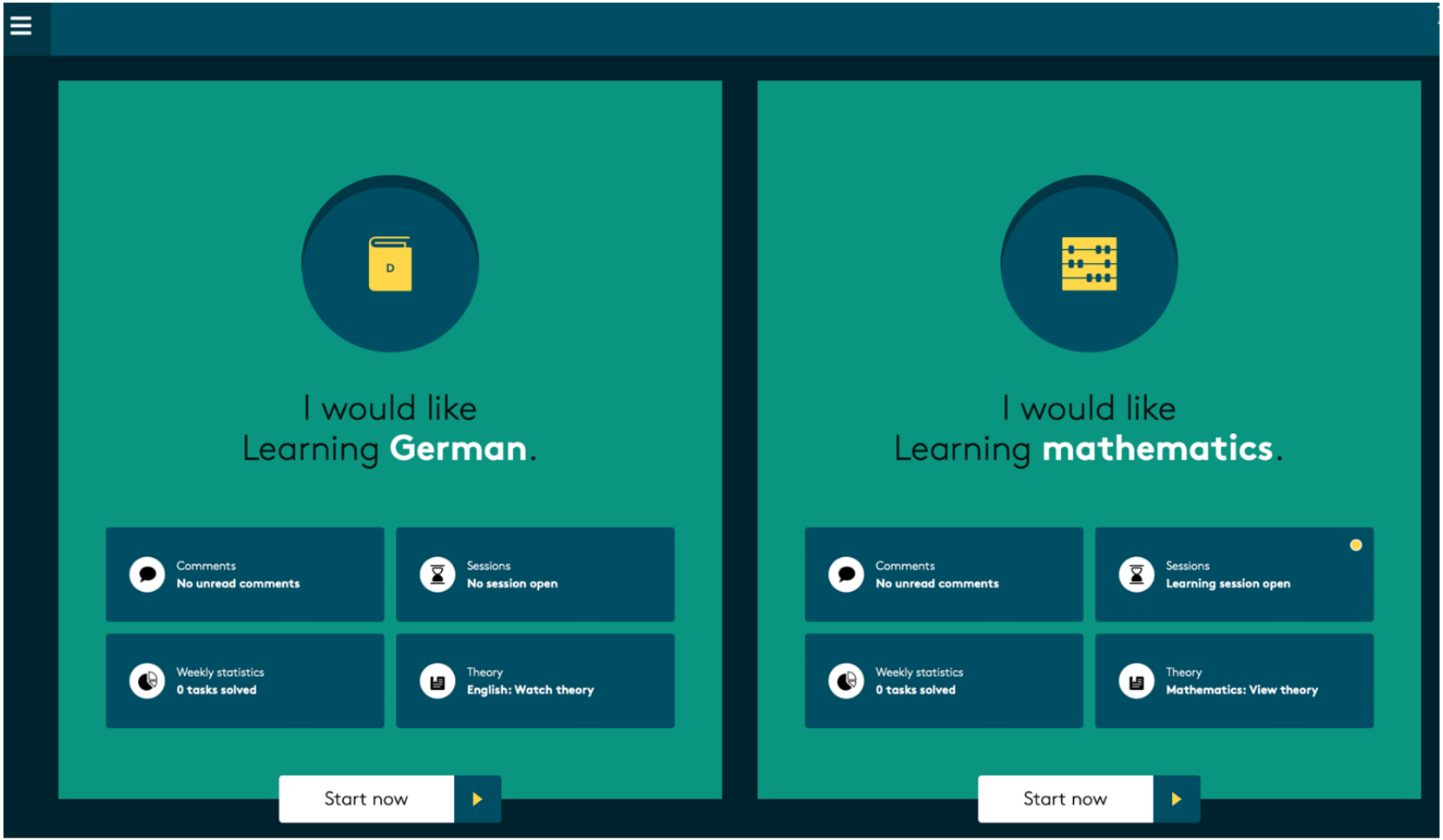


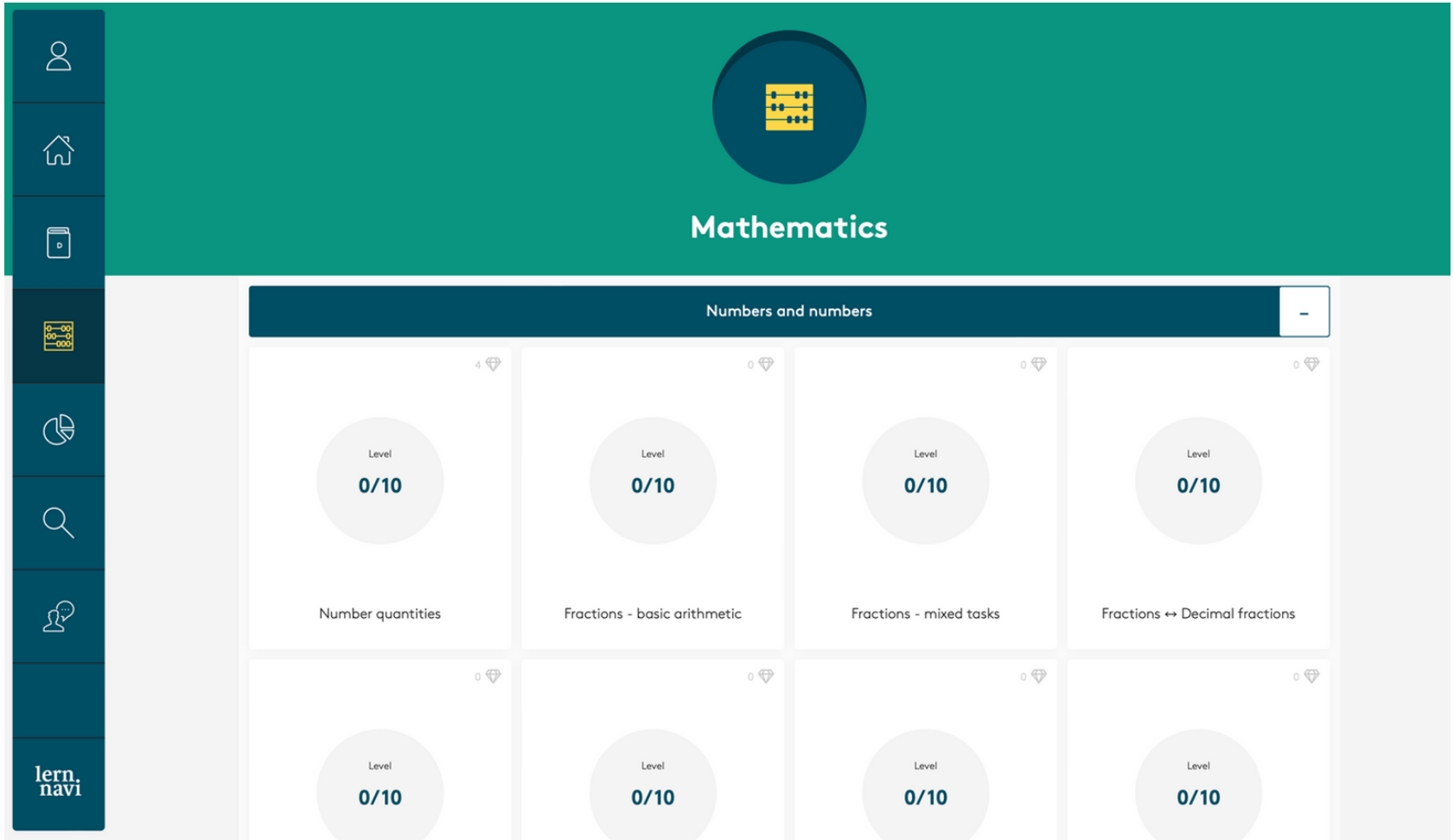
https://www.lernnavi.ch/

German and mathematics



Lernnavi is an instrument for promoting part of the basic subject-related study skills in German and mathematics. The development teams created the reference framework for German and mathematics based on the appendix to the framework curriculum. The cantonal student councils and the VSDL (Association of Swiss German teachers) or the DMK (German-Swiss Mathematics Commission) and the core group for canon mathematics were able to comment on these in the consultation process. Thanks to this approach and the broad support, a good fit with the framework curriculum is ensured.


I would like
Learning German.
Comments
No unread comments
Sessions
No session open
Weekly statistics
0 tasks solved
Theory
English: Watch theory
Start now
I would like
Learning mathematics.
Comments
No unread comments
Sessions
Learning session open
Weekly statistics
0 tasks solved
Theory
Mathematics: View theory
Start now


Mathematics
Numbers and numbers
4
Level
0/10
Number quantities
0
Level
0/10
Fractions - basic arithmetic
0
Level
0/10
Fractions - mixed tasks
0
Level
0/10
Fractions ↔ Decimal fractions
0
Level
0/10
0
Level
0/10
0
Level
0/10
0
Level
0/10
lern. navi

Learning:
Number sets

Arrange the set of numbers so that the lower set is always a subset of the upper set.

$\mathbb{Q}$ $\mathbb{N}$ $\mathbb{R}$ $\mathbb{Z}$

0 Comments

skip

nd numbers

Number quantities

Rule

Payment quantities

$\mathbb{N} = \{1, 2, 3, 4, 5, ...\}$= set of natural numbers

$\mathbb{N}_0 = \{0, 1, 2, 3, 4, 5, ...\}$= Set of natural numbers with zero

$\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$= Amount of integers

$\mathbb{Q} = \{\frac{p}{q} | p \ominus \mathbb{Z}, q \ominus \mathbb{N}\}$= set of rational numbers = set of fractions (quantities)

$\mathbb{R}$= Set of all real numbers (decimal numbers)

The following applies:

$\mathbb{N} \,.\, \mathbb{Z} \,.\, \mathbb{Q} \,.\, \mathbb{R}$

Lernnavi Data Overview

- Activity logs of users (how much time they spent per question, which topics they explored, whether they visited the theory pages, etc)
- Textual questions and answers given per students (as well as detail whether it was correct, incorrect or partially correct).

Name	Description
users	Demographic information of users.
events	Events done by the users in the platform.
transactions	Some events (like SUBMIT_ANSWER or REVIEW_TASK) have associated outcomes (e.g., the solution and the evaluation), In the transaction table, we have fine-grained information on the questions (called documents in lernnavi) answered. Each transaction has an associated a transaction token that can be mapped to the events table.
documents	The actual questions/problems. It is relevant to see the topics. The documents represent the questions and the textual pages provided to students.
topics_translated	The topics represent the taxonomy of categories shown in the Deutsch and Math dashboard. The name column corresponds to what the platform shows at https://app.lernnavi.ch/dashboard/1. The topics are organized in two levels, e.g., Orthografie > Rechtschreibprinzipien.
topic_trees	The topic hierarchy
feedback_history	The feedback refers to the hints or responses given to the student once they answered a task

Demo

- Demo account
  - mlbd@taskbase.com
  - ZzW&7x9W*%Ht99u@rjx#
- https://app.lernnavi.ch/home

Lernnavi | Project Ideas

- Time series behavioral analysis of students.
- Next answer prediction using students' previous interactions.
- Is there a link between task difficulty and how well users perform?
- Can we find different types of learners based on how they interact with the platform? What are these types?
- Do certain feature usage patterns (theory access, forum etc.) reveal unique user needs or behaviors? Does using a feature more often help users perform better, or worse?

Milestone M1

https://go.epfl.ch/mlbd-m1-2026

Fill out with team and start-up preference

Deadline: Tuesday, Feb 25th, 23:59