



Utrecht
University

TU/e EINDHOVEN
UNIVERSITY OF
TECHNOLOGY

19 July, 2025

The Educational Proof Assistant Waterproof in an Introductory Proof Course

Proof Construction and Learning Processes

**Pim Otte, Rogier Bos, Johan Commelin,
Jim Portegies**
ThEdu 2026

Outline

Motivation

Context

Methods

Results

Conclusion

What is Waterproof?

- Students struggle with writing proofs.

What is Waterproof?

- Students struggle with writing proofs.
- A proof assistant is software that helps you write proofs.

What is Waterproof?

- Students struggle with writing proofs.
- A proof assistant is software that helps you write proofs.
- Waterproof is an educational proof assistant using controlled natural language.

What is Waterproof?

- Students struggle with writing proofs.
- A proof assistant is software that helps you write proofs.
- Waterproof is an educational proof assistant using controlled natural language.
- Why?
 - Goal: To teach writing paper proofs.

What is Waterproof?

- Students struggle with writing proofs.
- A proof assistant is software that helps you write proofs.
- Waterproof is an educational proof assistant using controlled natural language.
- Why?
 - Goal: To teach writing paper proofs.
 - Immediate feedback.

What is Waterproof?

- Students struggle with writing proofs.
- A proof assistant is software that helps you write proofs.
- Waterproof is an educational proof assistant using controlled natural language.
- Why?
 - Goal: To teach writing paper proofs.
 - Immediate feedback.
 - Transferable to paper.

Context

- Developed in Eindhoven by Jelle Wemmenhove, Jim Portegies and many others [Wemmenhove et al., 2024].

Context

- Developed in Eindhoven by Jelle Wemmenhove, Jim Portegies and many others [Wemmenhove et al., 2024].
- Used in teaching Analysis 1 (since 2019, obligatory since last academic year).

Context

- Developed in Eindhoven by Jelle Wemmenhove, Jim Portegies and many others [Wemmenhove et al., 2024].
- Used in teaching Analysis 1 (since 2019, obligatory since last academic year).
- This talk: Evaluation at Utrecht University in “Bewijzen in de Wiskunde” [Otte et al., 2026]

Context

- Developed in Eindhoven by Jelle Wemmenhove, Jim Portegies and many others [Wemmenhove et al., 2024].
- Used in teaching Analysis 1 (since 2019, obligatory since last academic year).
- This talk: Evaluation at Utrecht University in “Bewijzen in de Wiskunde” [Otte et al., 2026]
- Related: Mizar, Isar, Lurch, Naproche, Diproche, Verbose Lean, Yalep, Proofweb, ProofBuddy, Deaduction, Edukera, ...

Related educational work

- Learning about Proof with the Theorem Prover LEAN: the Abundant Numbers Task [Thoma and Iannone, 2022]
- Comparative analysis of student proof structure following the use of Waterproof [Wemmenhove et al., 2026].

Demo

`https://impermeable.github.io/aquarium/`

Research questions

1. How does the use of an educational proof assistant in an introductory proof course affect students' ability to construct high-quality mathematical proofs?

Research questions

1. How does the use of an educational proof assistant in an introductory proof course affect students' ability to construct high-quality mathematical proofs?
2. How does the use of an educational proof assistant in an introductory proof course affect the students' learning process?

Research questions

1. How does the use of an educational proof assistant in an introductory proof course affect students' ability to construct high-quality mathematical proofs?
2. How does the use of an educational proof assistant in an introductory proof course affect the students' learning process? (see paper)

Methods

1. 2 out of 7 groups offered Waterproof (voluntary).

Methods

1. 2 out of 7 groups offered Waterproof (voluntary).
2. Research participation was opt-in.

Methods

1. 2 out of 7 groups offered Waterproof (voluntary).
2. Research participation was opt-in.
3. Surveys, observation, proof structure analysis, grades.

Participants

Study Programme	Waterproof	Control
Mathematics (no double bachelor)	7	14
Mathematics and applications	0	9
Mathematics/Computer Science	13	0
Mathematics/Economics	1	5
Mathematics/Physics	9	14
Other:	4	4

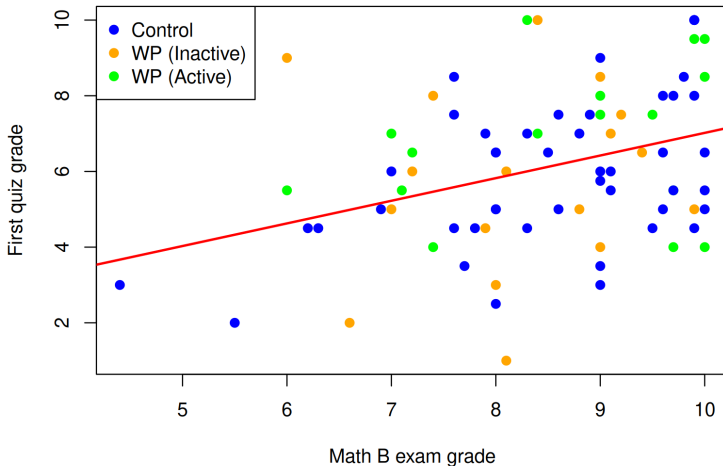
Performance on first quiz

Table: Mean Quiz 1 results by study programme and condition

Study Programme	WP (active)	WP (inactive)	Control
Mathematics (no double bachelor)	6.12	5.75	5.96
Mathematics and applications			4.61
Mathematics/Computer Science	8.00	6.78	
Mathematics/Economics	5.50		6.65
Mathematics/Physics	7.20	4.12	6.75
Other:	5.75	4.50	5.62

Performance vs Math B grade

Performance on first quiz compared to secondary school exam



Proof analysis

Table: Part of analyzed proof of participant 11 (no indentation)

Proof step	Code
...	...
- Case $q(14)$.	signpost-case
Assume that $q(14)$.	continued
Choose $y = 14$.	$\exists$ -intro
{Indeed, $14 \in \mathbb{R}$.}	domain-check
It holds that $\exists y \in \mathbb{R}, q(y)$.	continued
We conclude that $(\neg p(x)) \vee (\exists y \in \mathbb{R}, q(y))$.	signpost-concl
- Case $\neg p(x)$.	signpost-case
Assume that $\neg p(x)$.	continued
It holds that $\neg p(x)$.	continued
We conclude that $(\neg p(x)) \vee (\exists y \in \mathbb{R}, q(y))$.	or-intro,signpost-concl
End of proof.	continued

Difference in proof steps

Table: Average occurrences of selected logical steps per student, by group

Code	Control	WP (Inactive)	WP (Active)
or-elim	0.70	0.65	1.44
signpost-case	2.78	3.12	4.31
exists-intro	1.46	1.41	2.31

Transfer across language

Te bewijzen is dat $\forall y \in f(A), \exists x \in A, y = g(x)$.

Neem $y \in f(A)$.

$f(A) = \{f(x) \mid x \in A\}$, dus $\exists x' \in A, y = f(x')$.

Verkrijg deze $x' \in A$.

Er geldt dus $y \in f(x')$.

We moeten bewijzen dat $\exists x \in A, y = g(x)$.

Kies $x := x'$.

(Inderdaad, $x \in A$.)

We moeten bewijzen dat $y = g(x')$.

Per definitie van g geldt dat $g(x') = f(x')$,

dus het is waar dat $y = g(x')$.

We concluderen dat g surjectief is.

Table: Dutch proof of surjectivity of a function given on the exam by an active Waterproof user, showing a Waterproof-inspired proof

Selected survey responses

- “It helped me structure my answers, it taught me the basics on how to tackle problems, but it stopped me often when i tried to do something creative, because it was quite strict.”
- “The ‘by magic’ command is really fun”
- “While I do think Waterproof has definitely helped me with more quickly developing my proving skills, the actual program itself is still a bit buggy/slow”
- “I think it’s an excellent tool for learning proofs in mathematics, however I do think it needs to be accompanied by a textbook or lessons alongside it”

Future work

- Keep improving Waterproof (automation, exercises)

Future work

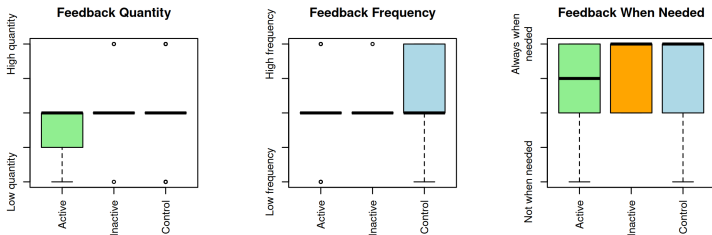
- Keep improving Waterproof (automation, exercises)
- Qualitative evaluation

Conclusion

- Waterproof helps students to make their proofs more explicit.
- Waterproof might be extra helpful for Computer Science double majors.

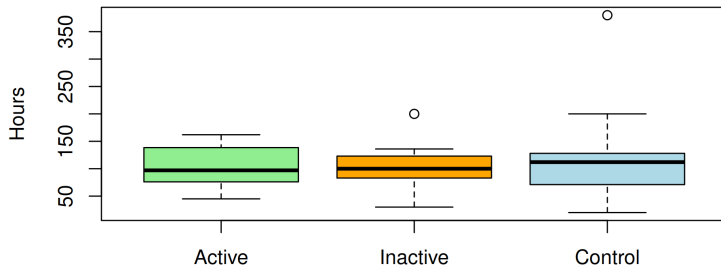
Thanks for your attention!

Perceptions about feedback



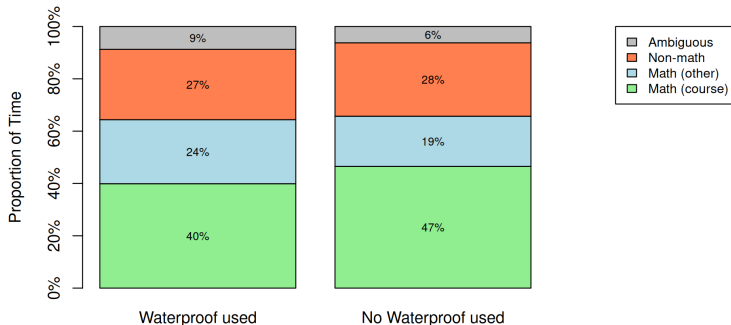
Survey results on perceptions about feedback

Time spent on the course



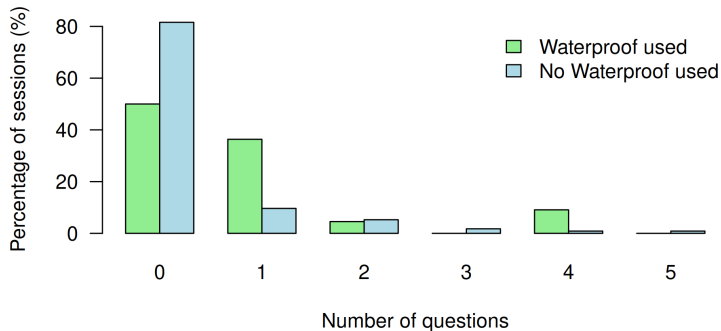
Survey results on time spent on the course

Time on task (in class)



Proportion of observed in-class time spent in each activity category, averaged over sessions and grouped by whether Waterproof was used.

Help-seeking behavior



Observations on number of questions asked by students

Waterproof screenshot

The screenshot displays the Waterproof proof editor interface. The top bar shows the file name 'example_waterproof.mv 2, U' and a 'Goals' tab. Below the top bar, the breadcrumb 'Real numbers > example_waterproof.mv' is visible, followed by a 'File verified' status bar. A message 'Import libraries (click to open/close)' is shown. The main area is titled 'Exercise 1' and contains a proof script starting with '27 Lemma exercise_1 : 2 is the infimum of [2, 5).', '28 Proof.', and a series of steps including '32 It suffices to show that', '33 2 is a _lower bound_ for [2, 5) ^', '34 $\forall \epsilon > 0, \exists a \in [2, 5), a < 2 + \epsilon.$ ', '35 We show both statements.', '36 - We need to show that $\forall c \in [2, 5), 2 \leq c.$ ', '37 Take $c \in [2, 5).$ ', '38 We conclude that $2 \leq c.$ ', '39 - We need to show that', '40 $\forall \epsilon > 0, \exists a \in [2, 5), a < 2 + \epsilon.$ ', '41 Take $\epsilon > 0.$ ', '42 Choose $a := 2.$ ', '43 { Indeed, $a \in [2, 5).$ }', '44 We conclude that $a < 2 + \epsilon.$ ', and '48 Qed.'.

On the right side, the 'Goals' tab is active, showing the goal 'We need to show' with a reference to 'example_waterproof.mv:42:17'. Below the goal, instructions are provided: 'Add the following line to the proof:', '{ Indeed, $a \in [2, 5).$ }', 'or write:', '{ We need to verify that $a \in [2, 5).$ }', and 'if intermediary proof steps are required.' Two buttons, 'Adopt suggestion' and 'Adopt suggestion 2', are visible. At the bottom of the goals panel, it says 'Afterwards, we need to complete other subproofs' with a reference to 'example_waterproof.mv:42:17'.

A Waterproof proof and example of feedback.

References I

Pim Otte, Rogier Bos, Johan Commelin, and Jim Portegies. The educational proof assistant waterproof in an introductory proof course: Proof construction and learning processes, 2026. URL <https://arxiv.org/abs/2606.26809>.

Athina Thoma and Paola Iannone. Learning about proof with the theorem prover lean: the abundant numbers task. *International Journal of Research in Undergraduate Mathematics Education*, 8(1):64–93, Apr 2022. ISSN 2198-9753. doi: 10.1007/s40753-021-00140-1. URL <https://doi.org/10.1007/s40753-021-00140-1>.

References II

Jelle Wemmenhove, Dick Arends, Thijs Beurskens, Maitreyee Bhaid, Sean McCarren, Jan Moraal, Diego Rivera Garrido, David Tuin, Malcolm Vassallo, Pieter Wils, and Jim Portegies. Waterproof: Educational software for learning how to write mathematical proofs. In Julien Narboux, Walther Neuper, and Pedro Quaresma, editors, *Proceedings 12th International Workshop on Theorem proving components for Educational software. (ThEdu'23)*, 2024.

Jelle Wemmenhove, Nick Hoofd, Alexander Schüler-Meyer, and Jim Portegies. Comparative analysis of student proof structure following the use of waterproof. To appear in post-proceedings of ThEdu 25, 2026.



Utrecht
University

Sharing science,
shaping tomorrow

TU/e LINDHOVEN
UNIVERSITY OF
TECHNOLOGY



pim.otte.dev