Supplementary Appendix A. Intervention Syllabus/Content and Task Map

Module overview

- Title: Assistive technologies for teaching DHH school students in computer science lessons
- Duration: 14 weeks (1 session/week, 2 hours)
- Format: blended (in-person + Google Colab practicum)
- Target group: preservice computer science teachers (Bachelor's years 3–4; Master's years 1–2)

Table A1. Module map (14 weeks)

Week	Topic	Activity	Product	Assessment criterion
1	DHH and access barriers in computer science lessons	Lecture; video lesson analysis	Reflective essay (300 words)	Accuracy of barrier identification
2	AT: ASR, captions, visual supports	Tool demonstrations	AT comparison table	Completeness; accuracy
3	WCAG 2.1 accessibility principles	Practicum: webpage audit	Audit report (1 page)	WCAG alignment
4	Generative AI (ChatGPT, Gemini): affordances and limits	Prompt testing	5 prompts + quality analysis	Critical evaluation
5	Designing accessible practicum instructions	Task editing	Adapted instruction	Clarity; visualization
6	Google Colab: building an accessible notebook	Practicum: first Colab	Colab draft (2 tasks)	Comments; visual cells
7	Visualizing algorithms (flowcharts, diagrams)	Creating visual supports	Set of diagrams (3)	Accuracy; accessibility
8	Teaching without oral instructions	Lesson simulation (role play)	Lesson script	Realism; completeness
9	Video cases: DHH classroom communication	Documentary materials review	Field notes	Reflective depth
10	Feedback and assessment for DHH	Rubric design	Task assessment rubric	Transparency; accessibility
11	Integrating AI into accessible materials	Colab refinement with AI	Interim Colab version	AI use + verification
12	Case study: resolving a didactic situation	Case analysis	Case study (2 pages)	Pedagogical reasoning
13	Finalizing artifacts	Peer review; editing	Final Colab + case	Rubric fit; quality
14	Project presentation and defense	Presentation (10 min)	Slides + Q&A	Communication; justification

Note: Week 9 originally included an observational visit to an organization working with people with hearing loss. In the final implementation, for some cohorts, this element was replaced by video-case/documentary analysis for training purposes, without collecting personal data.

Table A2. Key assistive technologies used in the module

Technology	Function	Use in the module
Automatic captioning (ASR)	Real-time speech-to-text	Lecture demo; accuracy analysis
Google Colab	Interactive coding environment with visual cells	Core tool for building accessible learning materials
ChatGPT / Gemini / Copilot	Generating/editing/rephrasing instructions and explanations	Faster drafting with mandatory verification
WCAG 2.1	Web content accessibility standard	Digital audits; compliance checks
Visual diagrams (flowcharts, UML)	Algorithm/process visualization	Replacing oral explanations with visual supports
Multimedia instructions (captioned video)	Learning videos with accurate captions and visual inserts	Creating and evaluating video-instruction quality

Table A3. Intervention fidelity and threshold attainment (EG)

Indicator	*n*	%
Completed ChatGPT/Gemini module (of included)	189/195	96.9
Developed a Colab notebook (of included)	187/195	95.9
Developed a case study (of included)	178/195	91.3
Submitted final artifacts for expert review (of included)	183/195	93.8
Colab ≥ 8/12 (Acceptable) (of rated)	181/183	98.9
Colab ≥ 10/12 (Optimal) (of rated)	171/183	93.4

Note: Participation percentages are calculated from the included EG sample (N = 195). Colab threshold percentages are calculated from rated submissions (*n* = 183). Rubrics and level descriptors for the Colab project and case study are reported in Appendix C (Tables C1–C2). Interrater reliability parameters and the adjudication procedure are reported in Appendix D (Tables D1–D3).

Supplementary Appendix B. Professional Readiness Measurement Instrument

Instrument description

We assessed preservice computer science teachers' readiness to use assistive technologies (AT) when teaching DHH school students using an author-developed 25-item questionnaire (5-point Likert scale: 1 = strongly disagree; 5 = strongly agree).

The authors developed the questionnaire for this study. An expert panel (n = 3) confirmed content validity. A pilot test with 15 students indicated acceptable clarity; no major comprehension issues were detected.

Components

- Motivational–value (5 items): attitudes, inclusion acceptance, professional responsibility
- Cognitive (8 items): knowledge of AT, accessibility principles, WCAG, tools (ASR, captions, AI)
- Operational (8 items): skills in designing accessible materials, adapting instructions, using AT in lessons
- Reflective (4 items): ability to evaluate AT effectiveness, critique decisions, adapt practice

Scoring rules

- Overall readiness index: mean of all 25 items (1–5)
- Component scores: mean of items within the component

- Missing data: if ≤20% of a component's items are missing, compute the mean of available items; if >20%, set the component to NA
- Reverse-coded items: none

Table B1. Full list of questionnaire items

No.	Item	Component
1	I consider it important to ensure accessibility of computer science lessons for students with hearing loss.	Motivational–value
2	I am ready to adapt my teaching methods for DHH learners.	Motivational–value
3	I believe computer science teachers are responsible for the accessibility of learning materials.	Motivational–value
4	Using assistive technologies in DHH teaching is a professional duty of the teacher.	Motivational–value
5	I am interested in learning new AT to support DHH learners.	Motivational–value
6	I know which assistive technologies are suitable for teaching DHH students in computer science lessons.	Cognitive
7	I understand WCAG 2.1 principles and how to apply them to educational materials.	Cognitive
8	I know how automatic captioning (ASR) systems work.	Cognitive
9	I understand how visual instructions can compensate for audio information.	Cognitive
10	I know how to use generative AI (ChatGPT, Gemini) to create accessible materials.	Cognitive
11	I understand AT limitations and when pedagogical review is required.	Cognitive
12	I know how to evaluate the quality of an accessible digital resource.	Cognitive
13	I understand which barriers DHH learners face in computer science lessons.	Cognitive
14	I can create an accessible Colab notebook with visual instructions for DHH learners.	Operational
15	I can adapt written instructions for programming tasks for DHH learners.	Operational
16	I can use ChatGPT/Gemini to generate accessible code explanations.	Operational
17	I can organize computer science practicum work so that DHH learners do not depend on oral explanations.	Operational
18	I can check a digital resource for compliance with accessibility principles (WCAG).	Operational
19	I can design a computer science lesson with minimal audio load.	Operational
20	I can use visual supports (diagrams, schematics, flowcharts) to explain algorithms to DHH learners.	Operational
21	I can provide feedback to DHH learners without relying on an oral channel.	Operational
22	I can evaluate how effectively AT supported a DHH learner during a lesson.	Reflective
23	I can adjust AT use based on observation of the learning process.	Reflective
24	I can critically evaluate AI output quality and correct errors.	Reflective
25	I am ready to learn from my mistakes in AT use and improve my practice.	Reflective

Respondent instruction

Rate how well each statement reflects your current knowledge, skills, and attitudes on a 1–5 scale:
1 = Strongly disagree; 2 = Rather disagree; 3 = Neutral / unsure; 4 = Rather agree; 5 = Strongly agree.

Supplementary Appendix C. Artifact Scoring Rubrics (Colab Resource and Case Study)

Table C1. Colab notebook rubric (0–12 points)

Criterion	0 points	1 point	2 points	3 points	Max
1. AI use (ChatGPT/Gemini/Copilot)	No use	Used without verification (errors remain)	Used and partly verified	Used and fully verified (no errors)	3
2. Accessibility adaptation (DHH)	No adaptation	Partial (1–2 elements)	Strong (3–4 elements)	Full (visual supports, text, structure)	3
3. Accessibility principles compliance	Non-compliant	Minimal (1 criterion)	Partial (2+ criteria)	Full compliance with key criteria*	2
4. Code comments	None	Minimal (<30%)	Partial (50–70%)	Full (all lines explained)	2
5. Practice tasks	None	1 correct task	2 correct tasks	3+ correct, testable, accessible tasks	2
TOTAL					12

Note: Key accessibility criteria in the module: text/background contrast, font size, text structure/readability, and basic WCAG 2.1 Level A requirements. During rater training, intermediate descriptors (levels 1–2 in some 0–3 criteria) were merged for calibration; the final 0–3 scoring scale remained unchanged. Thresholds: Acceptable ≥ 8/12 (66.7%); Optimal ≥ 10/12 (83.3%)

Table C2. Case study rubric (0–12 points)

Criterion	0 points	1 point	2 points	3 points	4 points	Max
1. Pedagogical reasoning	None	Superficial	Partly justified	Well justified	Full, coherent, evidence-based	4
2. Use of AI-based solutions	None	Mentioned, not explained	Use described	Use + quality check	Integrated with critique	4
3. Adaptability analysis	None	Descriptive	Partial (1–2 aspects)	Deep (barriers + solutions)	Full (barriers + solutions + evaluation)	4
TOTAL						12

Note: Thresholds: Acceptable ≥ 8/12 (66.7%); Optimal ≥ 10/12 (83.3%)

Supplementary Appendix D. Supplementary Materials on Reliability and Expert Scoring

Table D1. Expert panel composition

Expert	Role	Experience	Qualification
Expert 1	Department instructor (AI & IT)	12 years	PhD (Education), specialization: computer science teaching methodology
Expert 2	Department instructor (AI & IT)	8 years	Candidate of Pedagogical Sciences; experience designing digital learning resources
Expert 3	Practicing computer science teacher (partner school)	15 years	Highest category; experience with DHH learners (2 years)

Note: Experts received rubric training in a 2-hour workshop and completed calibration on five pilot submissions.

Table D2. Expert rating procedure

Stage	Description
1. Preparation	All submissions were anonymized and coded (ID 001–255).
2. Rater training	2-hour workshop: rubric explanation, examples, threshold discussion.
3. Calibration	Experts independently rated five pilot submissions; discrepancies were discussed.
4. Double scoring	20% of submissions (37 Colab + 35 case studies) were double-scored independently.
5. ICC computation	ICC(2,1) computed on independent ratings before reconciliation.
6. Adjudication	If discrepancy > 2 points, ratings were discussed to reach consensus.
7. Final scoring	Analyses used consensus scores.

Note: ICC(2,1) specification: two-way random, single measures, absolute agreement. ICC was computed before adjudication; analyses used consensus scores.

Table D3. Interrater agreement (ICC)

Artifact	ICC(2,1)	95% CI	Interpretation
Colab notebooks	0.87	[0.78, 0.93]	Excellent
Case studies	0.82	[0.71, 0.90]	Good

Note: *ICC interpretation:* <0.50 poor; 0.50–0.75 moderate; 0.75–0.90 good; >0.90 excellent. ICC > 0.80 indicates high rating reliability. Analyses were run in JASP 0.95.4.

Supplementary Appendix E. Codebook and Thematic Analysis Framework

Coding procedure

1. All interviews (n = 30) were transcribed verbatim.
2. Two researchers independently coded 30% of transcripts (9 interviews).
3. Discrepancies were discussed until consensus. Agreement levels were good/high.
4. All interviews were coded using the finalized scheme.
5. Codes were clustered into five themes.

Unit of analysis: meaning unit (a sentence or a sentence cluster expressing one idea)

Table E1. Interview codebook (codes → themes)

Code	Definition	Example (paraphrase)	Theme
Inclusion feasibility	Confidence that accessible computer science for DHH is possible	After converting instructions into text–visual form, the respondent saw accessibility as realistic	Theme 1
Attitude shift	Movement from doubt to acceptance of inclusive practice	The respondent moved from skepticism to believing DHH learners can program with proper organization	Theme 1
AI as a tool	AI as a way to speed up materials preparation	ChatGPT helped quickly rephrase instructions	Theme 2
Critical stance toward AI	Need to verify AI outputs	AI results must be checked for terminology and code accuracy	Theme 3
AI errors	Mention of specific AI mistakes (terms/code)	AI produced code with a syntax error	Theme 3
Pedagogical judgment	Teacher's role as final decision-maker	The teacher remains responsible for correctness and instructional appropriateness	Theme 3
Contact with DHH context	Experience of observing DHH communication	Videos/observations improved understanding of DHH needs	Theme 4
Empathy	Growth of empathy and professional responsibility	The respondent reported stronger responsibility for access	Theme 4
Transfer barrier: time	Lack of time to adapt materials	Concerns about workload limiting adaptation in schools	Theme 5
Transfer barrier: support	Lack of technical/methodological support	Absence of services/consultation infrastructure in schools	Theme 5
Transfer barrier: institutional	Administrative resistance / lack of resources	Potential barriers from administration or low prioritization of inclusion	Theme 5

Note: Quotes are paraphrased to protect confidentiality. All interview participants provided informed consent for anonymized data processing.

Table E2. Theme map (Theme → subthemes → meaning)

Theme	Subthemes	Meaning	Frequency (n/30)
Theme 1. Feasibility of inclusive computer science	Attitude shift; confidence in accessibility	From doubt to recognizing DHH learners can succeed with well-designed materials	29 (96.7%)
Theme 2. Practical value of generative AI	AI as a tool; faster preparation	AI reduces routine workload and speeds up accessible material creation	27 (90.0%)
Theme 3. AI limitations	Critical stance; AI errors; pedagogical judgment	AI requires verification; the teacher remains the final expert	22 (73.3%)
Theme 4. Value of exposure to DHH context	Contact; empathy; responsibility	Exposure strengthens empathy and professional responsibility	28 (93.3%)
Theme 5. Transfer barriers	Time; support; institutional	Transfer depends on school conditions, resources, and administrative support	26 (86.7%)

Table E3. Interview themes and prevalence (EG; n = 30)

Theme	n (%)	Core meaning
1. Feasibility of inclusive computer science	29 (96.7)	Accessibility is achievable through text–visual design of explanations and instructions
2. Practical usefulness of generative AI	27 (90.0)	AI speeds up materials design and reduces routine workload
3. AI limitations	22 (73.3)	Pedagogical review is required: checking terminology and code, editing outputs
4. Value of direct contact	28 (93.3)	Contact with the DHH community strengthens empathy and professional responsibility
5. Transfer barriers	26 (86.7)	Transfer depends on school conditions and the availability of support/resources

Note: Frequencies indicate the share of interviews in which the theme was explicitly coded; a single transcript could include multiple themes.

Supplementary Appendix F. Supplementary Tables on the Sample and Sensitivity Checks

Table F1. Demographic profile of participants by group (EG vs HCG)

Variable	HCG (*n* = 60)	EG (*n* = 195)	Test	*p*
Gender				
Women	35 (58.3%)	119 (61.0%)	$\chi^2(1) = 0.17$	0.681
Men	25 (41.7%)	76 (39.0%)	—	—
Age	M = 21.2 (SD = 1.8)	M = 21.4 (SD = 1.9)	t(253) = 0.69	0.489
Degree level				
Bachelor's	18 (30.0%)	82 (42.1%)	$\chi^2(1) = 3.28$	0.070
Master's	42 (70.0%)	113 (57.9%)	—	—
Prior experience with DHH				
Yes (any contact)	3 (5.0%)	8 (4.1%)	Fisher's exact	0.725
No	57 (95.0%)	187 (95.9%)	—	—

Note: DHH = deaf and hard of hearing. Differences in gender, age, and prior experience were not statistically significant ($p > 0.05$). Degree-level differences approached the threshold ($p = 0.070$) and were therefore examined in sensitivity checks (Tables F3–F4).

Table F2. Baseline equivalence (pre-test)

Readiness component	HCG (*n* = 60) M (SD)	EG (*n* = 195) M (SD)	*t*	df	*p*
Overall index	2.08 (0.79)	1.94 (0.81)	1.09	253	.278
Motivational–value	2.75 (0.48)	2.68 (0.45)	0.97	253	.332
Cognitive	1.48 (0.38)	1.42 (0.35)	1.05	253	.295
Operational	1.78 (0.42)	1.72 (0.40)	0.93	253	.352
Reflective	1.92 (0.40)	1.85 (0.38)	1.15	253	.252

Note. All baseline differences were non-significant ($p > 0.05$), supporting initial group equivalence.

Table F3. Subcomparison within EG: Bachelor's vs Master's (*n* = 195)

Variable	Bachelor's (*n* = 82) M (SD)	Master's (*n* = 113) M (SD)	*t*	df	*p*
Pre-test (overall index)	1.88 (0.78)	1.98 (0.83)	−0.86	192	0.391
Post-test (overall index)	4.12 (0.37)	4.22 (0.34)	−2.05	163	0.042
Colab artifact (0–12)	10.0 (1.6)	10.4 (1.4)	−1.87	159	0.063
Case study (0–12)	8.5 (2.3)	9.5 (1.9)	−3.42	149	<0.001

Note: Baseline differences were non-significant ($p = 0.391$). Master's students scored higher on case studies at post-test ($p < 0.001$), consistent with stronger pedagogical reflection. These differences do not alter the main conclusion about intervention effectiveness (ANCOVA adjusts for pre-test).

Table F4. Subcomparison within HCG: Bachelor's vs Master's (n = 60)

Variable	Bachelor's (*n* = 18) M (SD)	Master's (*n* = 42) M (SD)	*t*	df	*p*
Pre-test (overall index)	2.05 (0.82)	2.10 (0.78)	−0.21	27	0.836
Post-test (overall index)	2.88 (0.41)	2.98 (0.36)	−0.89	26	0.382

Note. In the HCG, Bachelor's vs Master's differences were non-significant at both pre- and post-test. Artifact scores (Colab/case study) were not collected in the HCG.