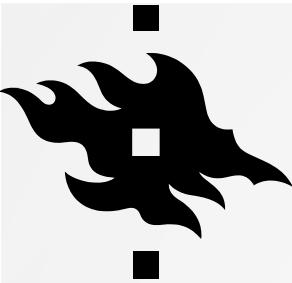


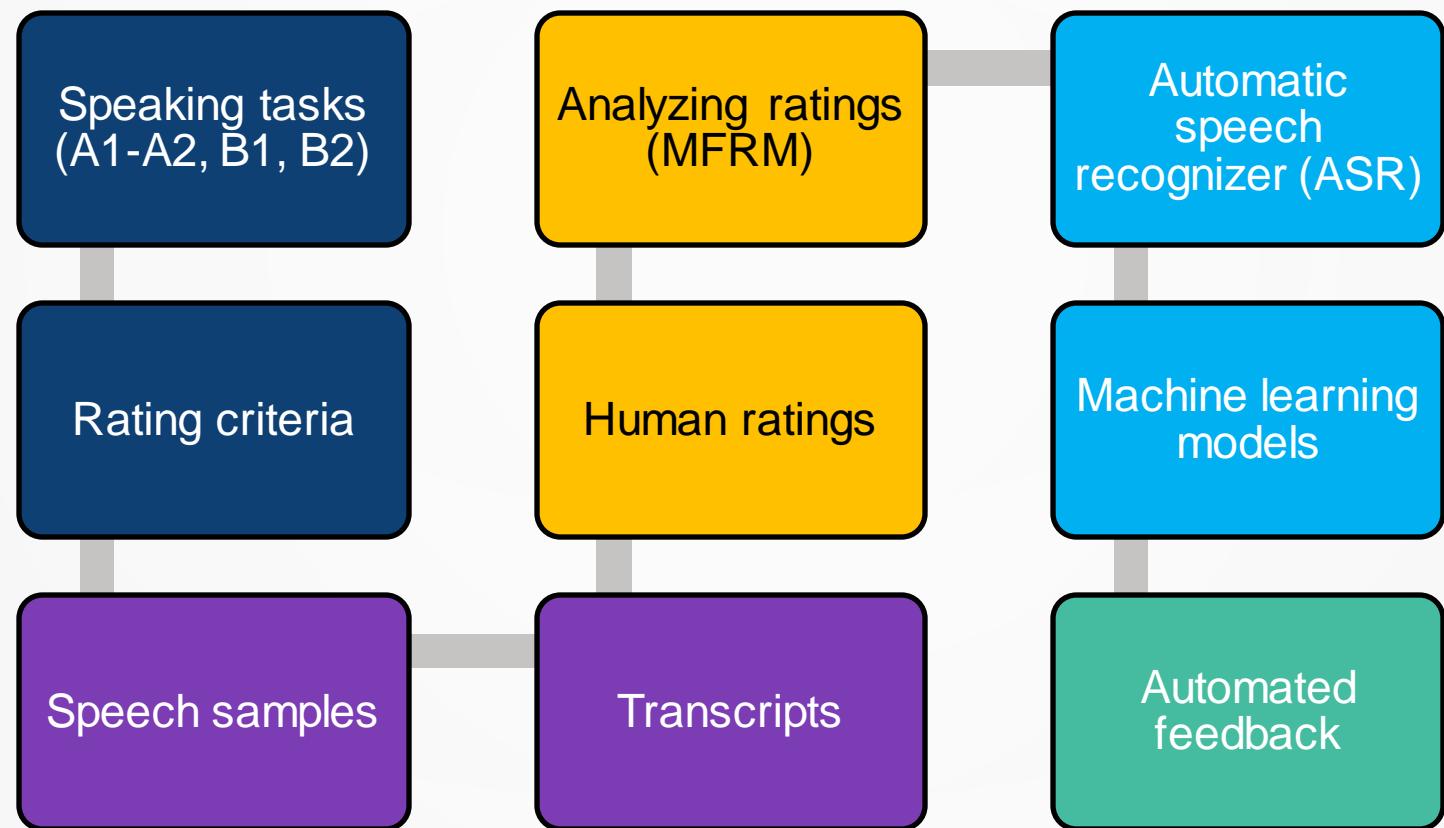


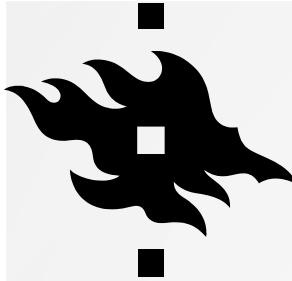
STAKEHOLDER BELIEFS: BIGGEST FEAR OR A DREAM COME TRUE?

DigiTala Closing Seminar | April 13, 2023
anna.vonzansen@helsinki.fi
raili.hilden@helsinki.fi



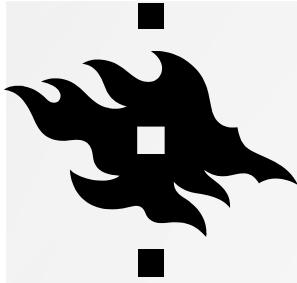
AUTOMATED L2 SPEAKING ASSESSMENT





SPEAKING TASKS

- Different test versions for proficiency levels (A1-A2, B1, B2)
 - Moodle (remotely / in class)
 - B1-B2: the curriculum for upper secondary education
 - A1-A2: beginners' university course
- Read-alouds, semi-structured, open-ended



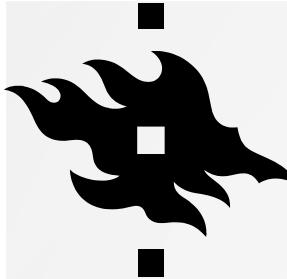
RATING CRITERIA

Holistic score

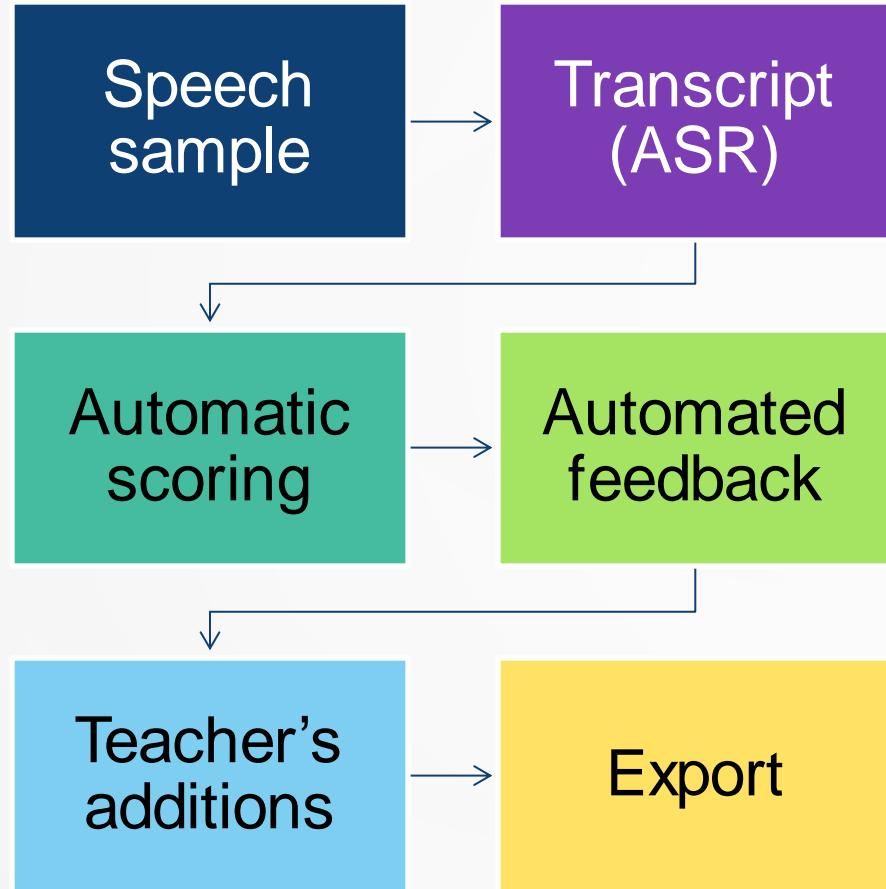
- Below A1-C2 (proficiency level)

Analytic scores(0–3/4)

- Task completion (does the speaker answer the question)
- Fluency (breaks and repetitions)
- Pronunciation (control of sound and prosodic features)
- Range (extent of vocabulary, structures and expressions)
- Accuracy (vocabulary and grammar errors > comprehensibility)



THE MOODLE PLUGIN



Evaluation report

Submitted: 10.05.2022 11.21:13

This feedback concerns only the speech sample you produced and it does not cover all aspects of your oral language skills. A machine produces your grades automatically. We have taught the machine with speech from other language learners together with other language-specific data.

There is no limit set for the number of attempts on this assignment.



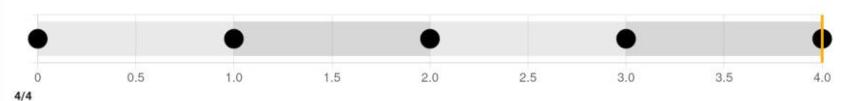
A transcript of your speech sample

öö mult oli jälj huppari teillä sinne kahviaan ellen

Analytic grading Proficiency level

Fluency

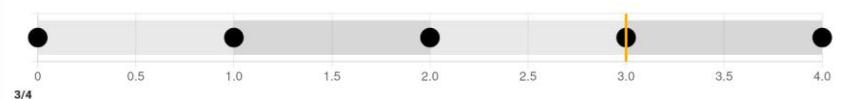
This measure reflects the speed, pauses, and hesitations in your speech.



Based on the automatic grading, it seems that your speech is very fluent and no disturbing pauses, breaks, or hesitations occur.

Pronunciation

Above you can see that the machine transformed your speech into text. There you can check whether you pronounced all the words right. This measure reflects how well the machine understands your speech. The speech samples that the machine has heard before affect its ability to understand you.



Based on the automatic grading, it seems that the machine understands you and there seems to be no major issues in your pronunciation.

Task completion

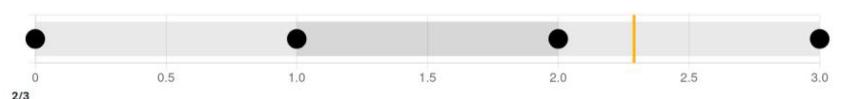
This measure is based on the previous responses that have been used in teaching the machine to grade this task.



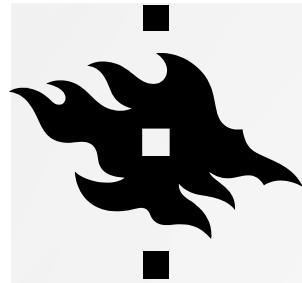
Based on the automatic grading, it seems that unfortunately, the machine has not heard this type of performance before and therefore failed to grade your speech.

Range

This measure reflects how much you have spoken and how comprehensive your vocabulary and sentence structures are.

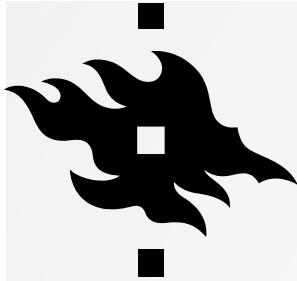


Based on the automatic grading, it seems that you use basic words and are able to form sentences.



STAKEHOLDER BELIEFS

Benefits	Challenges	Open issues
<ul style="list-style-type: none">• Tireless rater• Immediate feedback• Self-regulated learning• Saves time & money• Same criteria for all	<ul style="list-style-type: none">• Others listening• Others talking• Inexperience• Security, data protection	<p>Construct coverage? Reliable and fair? No prejudice? Test-anxiety?</p>



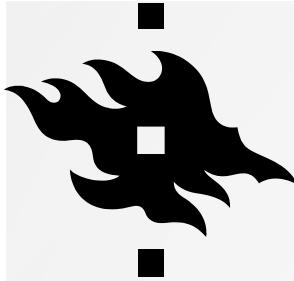
TEACHERS' FIRST IMPRESSIONS

RQ1 Possibilities and limitations for the use of the tool

- + a resource for self-regulated learning / teaching, part of the oral test / Kieliprofiili / homework, accessibility
- data protection, compatibility, negative washback
- ? transparency, reliability, customizability

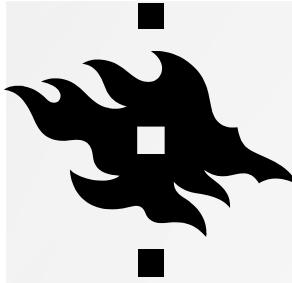
RQ2 Usefulness of the feedback

- + more feedback and speaking opportunities, transcript useful, support the use of CEFR level (they should use, but don't yet)
- analytic dimensions do not 100% match teachers' feedback, construct definition / coverage?, teacher needed to supplement and mediate the feedback to learners
- > Developing national rating criteria for assessing speaking



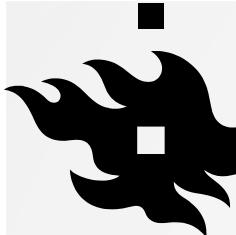
TEACHER PERCEPTIONS TODAY AND 10 YEARS AGO – ARE THE TIMES A-CHANGIN’?

- In 2012: 202 language teachers responded to an online survey
- In 2023: some of the questions were answered by 129 language teachers
- RQ1 What perceptions and experiences of oral language assessment construct and purpose do teachers report?
- Fluency, pronunciation and interaction skills considered as prior assessment criteria in both surveys
- Grammar less important
- Open-ended answers reflected the pertaining assets and dilemmas of speaking assessment in summative and high-stakes contexts



SHOULD THE ME LANGUAGE TESTS INCLUDE A SPEAKING COMPONENT?

		English A	Finnish / Swedish (B1)	Swedish / Finnish (native- level)	Swedish / Finnish (A)	German / French / Russian / Spanish (A)	B2 and B3 languages	Finnish (S2) / Swedish as a second language and literature
2012	N	218	208	202	203	206	207	-
	Yes	46,3% 01	33,2% 69	41,1 % 83	43,8 % 89	39,8 % 82	26,1 % 54	-
2023	N	127	124	121	124	123	124	121
HELSI HELSI UNIVERSITY OF HELSINKI	Yes	89 % 113	71 % 88	77,7 % 94	86,3 % 107	81,3 % 100	58,9 % 73	76,0 % 92

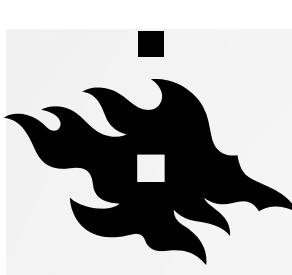


WHAT PERCEPTIONS AND EXPERIENCES OF AUTOMATED ASSESSMENT OF L2 SPEAKING DO TEACHERS HOST?

- Q10 How do you feel about automated assessment of oral skills?

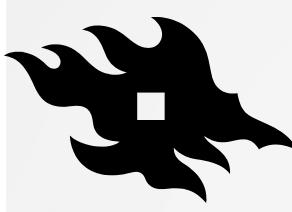
Question	Very negative	Quite negative	Neutral	Quite positive	Very positive	Mean	Standard deviation
Q10 n=129	1,6 % 2	11,6 % 15	36,4 % 47	34,1 % 44	16,3 % 21	3,52	0,953

Digitala project has produced the first tool to facilitate the practical implementation of introducing an oral part to the Finnish ME. The prototype is ready for use, but refining it for large-scale use requires additional resources for long-term maintenance.



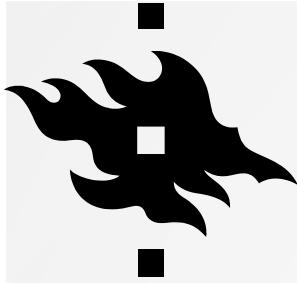
REFERENCES 1/2

- Alanen, T., Erkkilä, J., Harjunpää, T., & Heijala, M. (2022). Digitala Moodle plugin user manual (1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.6535377>
- ALTE (2020) Principles of Good Practice.
- Deeva, G., D. Bogdanova, E. Serral, M. Snoeck & J. De Weerdt 2021. A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education*, 162. <https://doi.org/10.1016/j.compedu.2020.104094>
- Evanini, K., & Zechner, K. (2020). Overview of automated speech scoring. In K. Zechner & K. Evanini (eds.), *Automated Speaking Assessment: Using Language Technologies to Score Spontaneous Speech*. Routledge, 3–20.
- Gu, L., & L. Davis. 2020. Providing SpeechRater feature performance as feedback on spoken responses. Teoksessa K. Zechner & K. Evanini (toim.) *Automated Speaking Assessment: Using language technologies to score spontaneous speech*. New York: Routledge, Taylor & Francis Group, 159–175.
- Kautonen, M. & von Zansen, A. (2020). DigiTala research project: Automatic speech recognition in assessing L2 speaking. *Kieli, koulutus ja yhteiskunta*, 11(4). Saatavilla: <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-kesakuu-2020/digitala-research-project-automatic-speech-recognition-in-assessing-l2-speaking>



REFERENCES 2/2

- von Zansen, A., Alanen, T., Al-Ghezi, R., Erkkilä, J., Harjunpää, T., Heijala, M., Kallio, H. (2022). DigiTala Moodle plugin. <https://github.com/aalto-speech/moodle-puheentunnistus>
- von Zansen, A. & Heijala, M. (accepted). Miten suomen ja ruotsin opettajat käyttäisivät puheen automaattiseen arviointiin kehitettyä työkalua? [How would Finnish and Swedish teachers use a tool developed for automated L2 speaking assessment?].
- von Zansen, A., & Hilden, R. (2022). "It was cool and comfortable!" Akateemisten alkeistason S2-opiskelijoiden kokemuksia tietokoneella suoritettavasta puhumisen kokeesta. teoksessa S. Routarinne, P. Heinonen, T. Kärki, A. Roiha, M-L. Rönkkö, & A. Korkeaniemi (Toimittajat), Ainedidaktiikka ajassa: Laajenevat oppimisympäristöt ja eri-ikäiset oppijat ["It was cool and comfortable!" Academic L2 Finnish learners' perceptions of a computer-based speaking test.] S. 72-90. Suomen Ainedidaktisen tutkimusseura. Turun yliopiston opettajankoulutuslaitos. <http://hdl.handle.net/10138/353562>
- von Zansen , A & Huhta , A 2022 , Developing Automated Feedback on Spoken Performance : Exploring the Functioning of Five Analytic Rating Scales Using Many-facet Rasch Measurement . in J. H. Jantunen, J. Kalja-Voima, M. Laukkarinen, A. Puupponen, M. Salonen, T. Saresma, J. Tarvainen & S. Ylönen (eds.) *Diversity of Methods and Materials in Digital Human Sciences: Proceedings of the Digital Research Data and Human Sciences DRDHum Conference 2022, December 1-3, Jyväskylä, Finland. University of Jyväskylä, Jyväskylä*, (pp. 211-229), Digital Research Data and Human Sciences (DRDHum), Jyväskylä, Finland, 01/12/2022 <http://urn.fi/URN:ISBN:978-951-39-9450-1>
- von Zansen, A., Kallio, H., Sneck, M., Kuronen, M., Huhta, A., Hilden, R. (2022). Ihmisarvioijien näkemyksiä suullisen kieltaidon automaattisesta arvioinnista, digitaalisesta arvointiprosessista sekä arvioitavista puheen ulottuvuuksista. [Human raters' perceptions of the automated assessment of oral language skills, the digital assessment process and the dimensions to be assessed from speaking performances.] AFinLA yearbook <https://doi.org/10.30661/afinlavk.114821>
- von Zansen, A., Sneck, M., Hilden, R. (2022). Lukioilaiden käsitykset ja heidän antamansa palautte suullisen kieltaidon automaattisesta arvioinnista. [Language learners' perceptions and their feedback on automated assessment of oral language skills.] LINGUAPEDA 2021. Conference Proceedings. Suomen ainedidaktisen tutkimusseuran julkaisuja. Ainedidaktisia tutkimuksia 21, p.188–219. <http://hdl.handle.net/10138/352128>



Moodle
plugin
(video)

Materials
(Zenodo)



helsinki.fi/en/digitala