Young students’ treatment of synthetic voicing as an interactional resource in digital writing

Niklas Norén, Helen Melander Bowden & Ann-Carita Evaldsson

To cite this article: Niklas Norén, Helen Melander Bowden & Ann-Carita Evaldsson (2021): Young students’ treatment of synthetic voicing as an interactional resource in digital writing, Classroom Discourse, DOI: 10.1080/19463014.2020.1814367

To link to this article: https://doi.org/10.1080/19463014.2020.1814367

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 14 Feb 2021.

Submit your article to this journal

Article views: 59

View related articles

View Crossmark data
Young students’ treatment of synthetic voicing as an interactional resource in digital writing

Niklas Norén, Helen Melander Bowden and Ann-Carita Evaldsson

Department of Education, Uppsala University, Uppsala, Sweden

ABSTRACT
This multimodal conversation analysis study is part of a larger video ethnographic project that explores the media literacy practices that children develop as they use digital and mobile technologies. The study investigates how Swedish students in grades 3–4 make use of text to speech (TTS) technology as an interactional resource during collaborative writing on iPads in the classroom. The results show that students routinely make use of synthetic voicings to display and claim knowledge about the voiced written units and negotiate writing roles with differing epistemic rights and obligations to assess voicings and practice repair.

KEYWORDS
Writing in interaction; conversation analysis; text to speech; epistemic stance; repair

1. Introduction
Young students increasingly learn to write using digital writing tools. Apart from engaging students in schoolwork (Dunn and Sweeney 2018), technical affordances (e.g. screens, visual spelling-correction) provide opportunities for talk and collaboration during writing activities (Čekaitė 2009; Musk 2016). The visual, embodied, and spoken aspects of writing in interaction have been fairly well explored and shown to construct a publicly shared writing process. Collaborative writing involves joint orientation to visual objects on screen (Greiffenhagen and Watson 2009) and concerted use of bodies and writing artefacts (Mondada and Svinhufvud 2016). Writing aloud, i.e. when writers simultaneously say what is written, makes the contents of the writing public, and holds on to the floor via demonstrating that work is still in progress (Komter 2006; Mortensen 2013). However, the interactional uses of automatic audible features in writing softwares are less fully explored, despite being increasingly used in educational settings in Sweden. Text-to-speech (TTS) technology in writing softwares builds on functions in the software/hardware that produce artificial voicings of the writing. The writing device activates a synthetic voice to automatically pronounce individual letters, and combines (synthesises) written units into words, sentences, and texts (see the methods section for more details). Synthetic voicings are designed to support individual writers’ on-line monitoring of their writing by providing instant audible feedback on letters, words, sentences and texts, in order to move the writing process forward. These functions are not obligatory and can be deactivated, turning the software into an ordinary word processor, but the idea is that the instant feedback enables writers to hear what they write on a systematic basis, which offers the
possibility for (especially beginner) writers themselves or persons in their surroundings to correct recognisable typing mistakes, misspellings, syntactic errors, and incoherence in the text (http://www.skolstil.se). TTS functions offer similar interactional affordances as writing aloud (Komter 2006; Mortensen 2013; Dalby Kristiansen 2017), such as making the writing process public, but also differ in terms of the basic structural organisation of the talk. But TTS functions also differ from other publicly available tools, such as visual marking of misspelled words on screen (Čekaitė 2009; Musk 2016), as no diagnostic functionality is offered apart from the opportunity to hear recognisable problems. These similarities and differences from writing aloud and visual diagnostic tools in writing softwares raise questions about how synthetic voicings are treated interactionally by the students.

Research on TTS functions has primarily focused on individual writers’ spelling or text-structuring processes that do not provide insights into how synthetic voicings are used interactionally. There are, to the best of our knowledge, no studies of TTS functions in digital collaborative writing. The knowledge of how TTS in digital writing devices works in the everyday classroom, and how students make interactional use of synthetic voicings when writing together, is therefore yet sparse. This study aims to increase our knowledge of students’ use of automatically produced audible representations of their writing during collaborative writing tasks. A multimodal conversation analytic approach is deployed to explore how Swedish students in the 3rd and 4th grades use software with TTS on iPads as an interactional resource when writing together. The main attention will be devoted to how synthetic voicings shape the students’ interaction. The affordances of synthetic voicings for corrections and repair during the writing process will be explored to some extent, but only when oriented to in the interaction between the students.

1.1. Previous studies on the role of synthetic voicing when using digital writing tools

There are only few studies on the role of synthetic voicing when using digital writing tools, and they are spread across different fields and methodologies.

Studies of synthetic voicing within the field of writing and learning how to write have mainly focused on its experimental effect on young writers’ spelling (e.g. MacArthur 1998) or its effect on the number of students’ keystrokes and sentences, amount of editing, story quality, degree of motivation and audience awareness (Borgh and Dickson 1992). Other studies have focused on its role as an aid for students’ text revising processes (MacArthur 2000; Garrison 2009), or as a tool to improve young students’ phonemic awareness and decoding skills (MacArthur et al. 2001). However, these studies mainly deploy experimental methodologies that do not provide access to everyday practices. In addition, these studies mainly focus on the impact of the technology on individual students’ linguistic knowledge and practical competences, and not on the interaction among students.

Research on writing in interaction is a growing field, however, and below we will present some studies within that field that are relevant for the current investigation.
1.2. Ethnomethodological and CA perspectives on writing in interaction

Studies of collaborative writing using ethnomethodology and conversation analysis (EMCA) (Mondada and Svinhufvud 2016) share a view on writing as multimodal and multiparty interactional processes, proposing ‘writing-in-interaction’ as a new field of research. Although including digital writing tools as part of the field, so far research has mainly worked with data involving writing by hand in various activity types, such as police interrogations (Van Charludorp 2014), interviews (Pälli and Lehtinen 2014), design meetings (Mortensen 2013), and local democracy meetings (Mondada 2016). Handwriting-in-interaction in classrooms is embedded within specific writing assignments that frame the activity, and CA studies demonstrate how text production is organised interactionally on a moment-by-moment basis. In such a vein, Szymanski (2003) investigates how text production in student peer groups is interactionally prompted and framed using grammatical resources, and how responses are designed to fit these prompts and frames as well as to project the next writing actions. Wigglesworth and Storch (2012) demonstrate that spoken feedback on handwriting in pairs or small groups provides systematic opportunities for language discussion and hence language learning. In a study of L2 students’ writing, Kunitz (2015) shows how a text emerges through various stages, where each stage is ratified interactionally by vocal response particles and embodied resources, and how these ratifying actions are crucial in the process of collaborative text shaping. Jakonen (2016) investigates the relation between individual and shared writing in the classroom and demonstrates how students negotiate access to each other’s writing tasks, using noticings and topicalizations to gain access, and orient to their writing as a shared activity. Other researchers have investigated handwriters’ practice of vocalising what is written while writing it. In such a vein, Mortensen (2013) shows that vocalising what is written on Post-its in a design group session displays the writing as a publicly available activity and creates a participation framework for the writing that includes the entire group, inviting the group to ratify, comment, correct, or make suggestions with relevance to the ongoing and projected writing. Dalby Kristiansen (2017) identifies a similar phenomenon in a project group meeting at a Danish university where the students are shown to use a particular writing aloud voice when formulating a piece of text in a work in progress. The writing aloud voice achieves a change in footing that displays the formulated text piece as well as any doubts and hesitations regarding its formulation and lays it out for scrutiny by the other group members.

These studies all demonstrate (hand)writing-in-interaction as multimodal and multiparty activities that are locally organised, and they situate the writing process within embodied, task-oriented, and institutional contexts. However, they do not give insights into the specific conditions of using digital writing technology in interaction as a collaborative activity.

1.3. Digital writing-in-interaction

Research on digital writing tools in collaborative schoolwork using EMCA perspectives has predominantly focused on participants’ repair practices based on visually available phenomena on screen (Greiffenhagen and Watson 2009; Čekaitė 2009; Musk 2016). Repair practices demonstrate what students notice during the writing process, and the type of
information they orient to when doing noticing. Greiffenhagen and Watson (2009) investigated how 13–14-year-old L1 students in England identified and corrected trouble sources only on the basis of their noticing of graphic and textual information on screen. The software did not mark objects on screen as trouble or possible trouble. Čekaitė (2009) investigated how 8th-grade L2 students in Sweden identified and corrected spelling errors on the basis of red underlining, i.e. the software’s diagnostic tool. Red underlining generated a publicly available ground for students’ actions, and the analysis showed that the marking provided interactional spaces for the students to jointly demonstrate an ability to find a relevant remedy. The students acted upon the underlining as a normative reference point of their repair activity. The student in control of the technology was also in control of desired authoritative forms of knowledge, which offered rights to evaluate the work of the group. Similarly, Musk (2016) investigated how 17-year-old students with Swedish as L1 studying English as a foreign language used red underlining while writing on a collaborative project. The students oriented to the computer as a repair agent when marking words with red underlining, but the students also identified visual trouble sources by saying the trouble word aloud, pointing at it, or explaining the problem. Red underlinings were often ignored, but when corrected, the correction was effectuated on the basis of the typist’s own epistemic access, without using the options provided by the software’s spelling function, thereby the progressivity of the typing process was maintained rather than slowed down by using the software’s digital affordances. Gardner and Levy (2010) also studied a collaborative multimodal web-based writing task in Australian schools, but within a more general focus on writers’ coordination of speech and use of the keyboard and mouse. The analysis showed that coordination was achieved by synchronising speech and writing on-screen (writing aloud), or by writing aloud while following the words on-screen with the mouse and while formatting the text with the mouse. Coordination was also achieved when the writer synchronised corrective activities on-screen with corrective talk by other participants, or when other participants synchronised the onset of talk with the projected ending of a written sentence. The data did not, however, involve any use of software diagnostics, such as red underlining, or of TTS functions.

A recent review of research on students’ writing on-screen (Kucirkova et al. 2019) points to a general lack of studies on the relation between children’s on-screen writing processes and social interaction. The research presented above provides valuable knowledge of handwriting and digital writing in interaction, but it is clear that the existing body of research within the field does not provide insights into the interactional aspects of students’ use of TTS representations when using digital writing tools. This study aims to address this research gap by investigating the role of students’ interaction when using digital writing tools with TTS technology.

1.4. Aims and research questions

The aim of the present study is to explore how Swedish 3–4th-graders use synthetic voicing function when writing together using the software Skolstil 2© on iPads. More specifically, we investigate how synthetic voicings contribute to organising the interaction between students, and how the students make use of synthetic voicings as resources for achieving collaborative writing and social interaction.
2. Materials and methods

2.1. Participants, writing assignments, and video recordings

The data are drawn from a larger video ethnographic study (Talking and texting on the move financed by Marcus and Amalia Wallenberg Foundation under Grant MAW 2014.0057) that explores the media literacy practices that children develop in interactions with peers as they engage with digital technologies in school, after school, and at home (appr. 150 hours of video recordings). The current study draws on fieldwork that was conducted in two Swedish primary schools: first, a grade 4 with children aged approximately 10 years (May 2016), predominantly from ethnically homogeneous middle-class family backgrounds, and second, a grade 3 with children aged approximately 9 years (October 2016) in a school located in a low-income multi-ethnic neighbourhood. A regional ethic committee has approved the study. Written consent was obtained prior to recordings from teachers and the children’s guardians. The children were informed about the research project and their rights to decide whether they wanted to participate or not during fieldwork. All participants were assigned pseudonyms to protect their identities.

The focus participants were four students in 3rd grade and four students in 4th grade. In each school, students engaged in digital collaborative writing during two lessons, which were selected for analysis. All students worked in pairs, sharing iPads or having one each. The writing assignments differed between the two schools.

The assignment in 3rd grade was to write texts together in pairs about different fruits that the teacher had brought to school. The fruit assignment lasted for two weeks, and during the start of each writing lesson, the assignment was posted on the white board to provide examples of themes to write about, such as shape, colour, origin, where it grows, etc. First one student dictated the contents and the other student wrote the text on the iPad, and then the participants shifted roles halfway through the lesson. Hence, each text was collaboratively written. The assignment in 4th grade was to write poems about Easter during two lessons right before the Easter break. The poem assignment had figurative language as a central theme, so each sentence about Easter would have to contain a metaphor or a parable. One of the students in the pair had already written a poem by hand, and now wrote it on the iPad. The other student in the pair wrote a poem by hand on a piece of paper. A lot of interaction occurred between the students during the writing, and the talk was mostly centred around the poem written on the iPad, partly because of the visual and audible affordances of the technology.

2.2. The writing software

The students’ collaborative writing was accomplished on iPads with the software Skolstil 2© (http://www.skolstil.se) with a TTS functionality that converts written letters, words, sentences, and texts into letter sounds and synthesised speech. TTS is offered as two separate systems with different voice qualities: a) letter sounds, with standard Swedish phonemic representations of individual letters (e.g. [k] and [a] for the letters ‘q’ and ‘a’); and b) speech synthesis (Acapela©, Elin) that combines different linguistic units into larger ones. Letter sounds are activated when a letter key is stroked. Words (or word-like units) are activated when using the space bar, sentences when using final punctuation (e.g. full stop, question mark, exclamation.
mark), and text units (all text from the beginning of the document) when the speaker icon above the writing surface is tapped. Swedish words are voiced in their standard forms, while misspelled or non-Swedish words are voiced according to Swedish phonics, which leads to deviant or mispronounced voicings. We use the term synthetic voicing when referring to the synthetic speech output produced by the iPad, as these synthetic utterances are voicings (or audible representations) of the students’ keyboard input.

The software features a spellchecker for written words that is connected to a vocabulary database. Deviant words are marked with grey colouring, and similar word forms in the database are offered within a black pop-up box. There is no grammar checker (parser) built into the software. The spellchecker and the TTS function are not interconnected. Hence, synthetic voicings only offer students the possibility of listening to how written units sound when pronounced as they are written. In these data, all synthetic voicings were made publicly available for everybody within listening range, as no headphones were used.

2.3. Analytic approach

Building on multimodal Conversation Analysis (CA) (Goodwin 2013; Heritage 2012b; Schegloff 2007), the analyses focus on the sequential organisation of actions that synthetic voicings emerge within, and on the epistemic organisation of students’ responses to voicings. The notion of adjacency pairs is of particular relevance to our study, as it describes a basic organisation of talk-in-interaction as constructed by first pair parts (FPP), i.e. initiatives that make a certain set of responses relevant, and second pair parts (SPP), i.e. responses that are made conditionally relevant by a particular FPP (Schegloff 2007). Generic examples of such adjacent pairing of actions are questions that make answering relevant or requests that make granting or rejecting relevant (Sidnell 2014). After a second pair part, the first point of possible sequence closure, participants have the possibility to confirm or acknowledge the response in a third turn receipt, or expand the sequence. Such post-expansions may, for example, be assessments that make further assessments relevant, explanations that account for some aspect of the action, or repair-initiatives that identify problems within the preceding sequence (Schegloff 2007). By analysing the sequential organisation of the students’ writing-in-interaction we demonstrate the orderliness of the students’ collaborative meaning-making when responding to synthetic voicings.

Another fundamental aspect of interaction concerns the role of epistemics for action formation and sequence organisation (Heritage 2012a, 2012b). In the context of educational settings, there is a growing body of research focusing on a variety of topics such as epistemic relations, negotiations and displays in teachers-student interactions (e.g. Heller 2017; Koole 2010; Kääntä 2014; Sert 2013), dynamic and changing epistemic stance taking in peer interaction (e.g. Melander 2012; Rusk, Pörn, and Sahlström 2016), the role of learning materials and embodied conduct for the establishment of epistemic relations (e.g. Jakonen 2015; Svähn and Melander Bowden 2019), etc. In our study, we analyse how participants in interaction monitor each other with respect to epistemic access or status, epistemic primacy, and epistemic responsibility (Stivers, Mondada, and Steensig 2011). We explore the epistemic organisation of the students’ responses to the synthetic voicings as they enact different writing roles with different rights and obligations to contribute to the writing process, depending on the epistemic status these roles assign to each student, and how such actions contribute to the action formation within and beyond the sequence.
The notion of participation frameworks foregrounds the interactive work of hearers and speakers as well as how the synthetic voicings figure in the organisation of actions (Goodwin 2013, 2018). We use Goffman’s (1981) analytic framework to conceptualise different production formats of an utterance in terms of animator, author and principal, in order to describe how a single strip of talk may contain an array of structurally different kinds of laminated speakers (see also Goodwin and Goodwin 2004). In the following analyses, animations merely work as sounding boxes for what someone else said or wrote. An author may be held partially responsible for how a spoken or written utterance is formulated, but may not be held fully responsible for the action. A principal may or may not be speaking or writing, but can be held fully responsible for the action. We will thus demonstrate how the children in interaction with each other and the software/synthetic voice, shift between positioning to give voice to their own and others’ actions.

The analyses are based on extracts from video-recordings and encompass the participants’ use of verbal, embodied, and material resources in situated interaction (Goodwin 2018). The analytic process, which started with an unmotivated exploration of the children’s interactions using Skolstil, came to focus on students’ immediate responses to synthetic voicings of words, sentences and text units (a text was formally defined as a unit exceeding one sentence). The total numbers of word and text voicings were equally distributed between the 3rd and 4th grade classrooms, while a larger number of sentence voicings were produced in the fruit assignment. In total, 119 occurrences of synthetic words, sentences and texts were found and analysed in the focus recordings (see Table 1). All synthetic voicings of words, sentences, and text units in the focus recordings were identified and transcribed within their immediate interactional context, following conventions developed within EMCA (Jefferson 2004; Mondada 2018, see Appendix for symbols used here).

Extracts were chosen on the basis of their illustrative quality and representativity to the entire collection, rather than number of occurrences. The basic organisation of keyboard input and synthetic voicings (extract 1) was consistent through all 64 synthesised words in the data. The students’ use of body, gaze and spoken responses to construct a shared accountability in the analysed sequential position (extract 2) occurred systematically after sentence voicings (about 13 occurrences) and the extract was chosen because it illustrated all of these resources. Extracts (1) and (2) were also chosen because together they illustrate a longer stretch of continuous interaction, compared to choosing two separate extracts. Initiations or repair in the interactional space following synthetic voicings were recurrent in the data (about 27 occurrences) and Extracts 3 and 4 were chosen because they involve different types of trouble sources (a grammatical error and a misspelling) and two types of outcomes of the students’ collaborative repair activity that illustrates both affordances and limits of the TTS tool when encountering trouble during collaborative writing.

Table 1. Synthesised writing units.

<table>
<thead>
<tr>
<th>Class and assignment</th>
<th>Student pairs</th>
<th>Synthesised units</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Word</td>
</tr>
<tr>
<td>3rd grade</td>
<td>Wanja and Sylvia</td>
<td>10</td>
</tr>
<tr>
<td>Fruit text</td>
<td>Laila and Martin</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td>33</td>
</tr>
<tr>
<td>4th grade</td>
<td>Alvin (iPad) and Bert</td>
<td>0</td>
</tr>
<tr>
<td>Easter poem</td>
<td>David (iPad) and Erik</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>64</td>
</tr>
</tbody>
</table>
3. Results

In the following extracts, we will demonstrate four main aspects of digital writing in interaction: firstly, how the writing of a word or sentence together with the activated synthetic voicing (*the voicing sequence*) is routinely used by the students as an audible diagnostic tool for the ongoing writing process (extract 1); secondly, that the students orient to the synthetic voicings of words, sentences, and text units as animating voices of their own writing (extracts 1–4); thirdly, how the voicings are embedded in the students’ interaction when treated as a resource for accounting for and correcting parts of the written products (extracts 2–4); and fourthly, how voicings are used as resources for negotiating production roles with different epistemic rights to assess voicings and effectuate repair (extracts 3–4).

3.1. The voicing sequence as a routine audible diagnostic tool

Extract (1) is drawn from a lesson where the students Sylvia and Wanja work on a collaborative text about fruits, and they have arrived at oranges. The two written words in the extract, *blir* (*will be*) and *den* (*it*), are first dictated by Sylvia (lines 1 and 6) and then typed on the iPad by Wanja (lines 2 and 7). These words are part of the emergent sentence *då blir den inte god* (*then it won’t be nice*), referring to the taste of oranges after using toothpaste (see extract 2). While typing the letters, the letter-voicing function in the software activates the phonemic sounds of each letter (lines 3 and 8). These letter voicings are normally not publicly noticed by the students, which is also the case in this extract. When the spacebar is used after each word, the synthetic voicing of the whole word is activated, and the speech synthesis generates the words (lines 5 and 10).

**Extract 1.** Assignment: Collaborative text about fruits: here oranges. Sylvia (S), Wanja (W), Letter voicing (LV), Synthetic voicing (SV).

1. S:       blir
           will be
2. W:       [b-l-i-r] SPACE       FPP
3. LV:      [”b l i r”]        FPP
4. W:       ((removes hand from screen, 0.2))
5. SV:      "blir"=
           will be
6. S:       =den
           it
7. W:       [d-e-n] SPACE       FPP
8. LV:      [”d e n”]       FPP
9. W:       [((#1 removes hand from screen, turns tow Stella))]
10. SV:     ["de[n"          SPP
           it
11. S:     inte
           not
This extract demonstrates the routine organisation of the interaction following synthetic voicings in terms of sequences, where the writing action (here individual words) together with the TTS activation (here the space bar) constitutes an initiation, and where the synthesised voicing constitutes a technology-mediated response. These contributions achieve a local sequence or meaning unit – the voicing sequence. We use the notions of FPP and SPP to account for these contributions’ sequential relation. Writing and activation of the speech synthesis (the FPP) project the software’s synthetic pronunciation of the writing into the next position, and the synthetic utterance (the SPP) animates, interprets, and makes the previous writing publicly available on a regular and projectable basis. Since conventional writing objects (space and punctuation) are assigned the function of activating the synthetic voicing, the students cannot avoid activating the synthetic voicing, unless they stop writing.

The students routinely orient to the synthetic voicing as an animation (Goffman 1981) of their own writing project, i.e. the voicing makes different aspects of their own writing skills and shortcomings publicly available for assessment. They do not seem to assess the technology’s ability to represent their writing (blaming the tool or treating it as an accountable participant) but seem to orient to it as an authoritative representation of their own writing. In this respect, the speech synthesis is used as an audible diagnostic tool, i.e. a digitally generated audible assessment of the correctness of the written form. This audible diagnostic tool is different from visual diagnostic tools, such as red underlining, which actively mark deviations from the linguistic vocabulary in the software, something the audible diagnostic tool in Skolstil does not do. During the synthetic voicings in the extract, the students listen to the animation of the written words, and then Sylvia immediately moves on to dictate the next word (lines 6 and 11). Wanja’s instant removal of her hand from the screen after activating the voicing (lines 4 and 9, Figure 1) displays publicly that her typing is done. The embodied move also makes interactional space for the speech synthesis to pronounce the written word and yields the next turn to Sylvia when Wanja turns to her. In terms of the analytical framework provided by Goffman (1981), Wanja’s embodied actions present her as the author of the sentence (here: the one who merely typed the words), while she attributes the role of principal to Sylvia (here: the one who gets to decide which words are correct). Thereby, Sylvia is attributed the epistemic status of knowing if the voicing is correct, which makes it relevant for her to assess the voicing in next position. By immediately dictating the next word, rather than initiating repair, Sylvia treats the embodied move as an invitation and displays an assessment of the voicing as correct. After both voicings, the students’ next action is latched to or is in slight overlap with the voicings, possibly in orientation to progressivity within the wider writing task. By moving on, and not confirming, assessing, or initiating repair, the students demonstrably treat the voicings as a recognisable, unproblematic, and correct animation of their writing.

In the analysed material, synthetic voicings routinely offer a sequential opportunity for the students to check and confirm that they have written the intended word, a word that fits with the contents of the ongoing text, and is formally correct. The
assessment of the voicing is routinely made on the basis of their own literacy knowledge (in a broad sense) of how the word should sound and how it should fit into the ongoing writing project, without recruiting help from the spellchecker in the software or from other persons. About half of all word-voicings and about a quarter of all sentence-voicings and whole-text-voicings in our data are tacitly, but publicly, treated as correct by immediately moving on, a frequency that confirms this practice as routinised, especially on the word level. The organisation of writing actions (including activations), synthetic voicings, and next actions in the extract illustrates a basic pattern within the data where the students routinely use the synthetic voicing as a support resource for linguistic forms they can predict and assess on the basis of their previous linguistic knowledge.

On the one hand, the students seem to mainly treat the voicings as animations of their own writing, thereby not treating the iPad as an agent that can be blamed or held accountable. In line with Greiffenhagen and Watson’s argument (2009), the device is not a participant. On the other hand, the voicings appear to construct assessment spaces on a structural level of talk that resembles SPPs in adjacency pairs, and this organisation attributes to the iPad (at least) some structural features that resemble a turn at talk. Therefore, in line with Luff, Gilbert, and Frohlich’s (1990) argument, the device could be viewed as a participant. It is, however, difficult to decide whether synthetic voicings are participants or just audible technological

Figure 1. The writer removes hand from screen and turns to the dictating student.
resources; they seem to be something in between, depending on how they are treated in the local interactional context.

3.2. Post-expansions – displays of intersubjectivity and accountability

Synthetic voicings also offer sequential opportunities to respond to the voicing sequence in various other ways, which extend and develop the sequence beyond a first possible completion point. By analogy with basic sequence organisation in mundane conversation (Schegloff 2007), these actions are analysed as post-expansions of the voicing sequence. The students’ actions in the position immediately after the voicing sequence demonstrate that an interactional space is created where the students may do different things before moving on to the next writing unit.

In the following extract (2), which is the continuation of extract (1), Wanja finishes the sentence with a full stop (line 6) and begins a new line on the screen with the return key. The software then generates the sentence (line 8). After the synthetic animation of the sentence is finished, Wanja does not move on to the next word, but turns towards the dictating student Sylvia, who returns the gaze (lines 10–11, Figure 2). The students nod to each other in lines 12–13, and then Sylvia says yes, treating Wanja’s gaze as a request for confirmation of the sentence.

Figure 2. The writer keeps hands away from screen and turns to the dictating student.

1. W: \[ i-n-t-e \] SPACE
2. LV: \[ “i n t e” \]
3. W: ((removes hands)) \[ ((turns tow S)) \]
4. SV: \[ “inte” \]
5. S: god (gazes at screen)
   nice (tasty)
6. W: \[ g-o-d \] STOP RETURN
7. LV: \[ “g o d” \]
   nice (tasty)
8. SV: ”den e lite sur när man borstar tender it is a bit sour when you brush your teeth
da blir den inte god.”
them it won’t be nice
9. W: -> ((turns tow S))
10. S: (#2 turns tow W)=
11. W: =((nods \[ turns back tow screen)) \]
12. S: \[ ((nods)) \]
    \[ ja. \]
yes
13. (0.8)
14. S: för att (.) när ja borstar mina tänder
   because when I brush my teeth
15. *dom e ( ) såhär.=
   they’re like this
16. *((brushing gesture))
17. W: a ja vet när man har åti frukt å äpple
   yeah I know when you’ve eaten fruits and apple
18. \[ vhua:]i:
   yuk
19. \[ ((covers mouth with hand)) \]
20. \[ mm::: \]
21. W: okej? ((turns tow assignment *and looks at it ))
22. S: *((turns tow assignment))
23. S: ehh:: (.) kärnorna.
   ehh:: the pips

By means of mutual gazes and nods, the participants co-produce an evaluation of the sentence voicing as correct and acceptable, and then Sylvia starts to account for the content of the sentence (that brushing your teeth after eating oranges makes them taste sour) (line 16). The account is left unfinished, and Wanja completes it collaboratively (Lerner 2004) in the next turn (line 19). They both accompany the collaboratively performed accounting with illustrative gestures in ways (lines 18 and 21) that upgrade the intersubjectivity that was previously displayed with nods and gazes. When she has finished accounting, Wanja closes the post-expansion sequence with the response
particle *okej* (line 23) and turns towards the assignment instructions on the white board, and then Sylvia initiates a new round of dictating and writing (line 25). The whole sequence from the noddings until they mutually orient to the next sentence belongs to the same post-expansion sequence, as the actions deal with the voiced sentence and establish its contents as acceptable for both parties.

### 3.3. Post-expansions – initiating repair and negotiating epistemic rights

The students also use synthetic voicings as an interactional resource for initiating and effectuating repair. Structurally, initiations of repair are analysed as post-expansions of the voicing sequence in a similar way to displays of intersubjectivity. A repair initiated in the sequential slot that follows after a voicing sequence expands that sequence beyond its first possible point of completion. The students’ trust in the synthetic animations of their co-authored writing may explain why repairs are recurrent in this sequential slot. In addition, these repair practices also demonstrate the students’ methods of using the voicing sequence as a resource to achieve and manage different participation roles, with different epistemic rights, in the writing activity.

In the next extract (3), David is writing a poem about Easter on the iPad. He has previously prepared the poem in handwriting on a piece of paper. Erik sits next to him, writing his own Easter poem by hand, preparing it for writing it on an iPad later on. When David finishes his sentence with a full stop (line 2), Erik overhears the following synthetic voicing of the sentence, and notices a problem regarding the indefinite article preceding the noun *träd* (tree).

**Extract 3.** Assignment: Individual poems about Easter. Synthetic voicing (SV). David (D) has moved on to write his poem on the iPad, and Erik (E) still writes on his poem by hand.

1. D: "påskljörö +sådär+
daffodils there
2. D: **STOP**
3. SV: "påskiljor är som en gult *träd."
daffodils are like a (COM) yellow tree
4. D: **VOICE**
5. E: -> ett [gult träd.
   a (NEUT) yellow tree
6. SV: ["påskens är som en godis dag."
   the Easter is like a candy day
7. D: (åhåhåhå:):(h):(h):(h))
   *((turns tow E))
8. SV: ["äggen är som en liten sol."
   the eggs are like a small sun
9. (1.0)
10. SV: "påskharen hoppar så långt."
    the Easter bunny jumps very far
11. (1.0)
12. SV: "påskiljor är som en gult träd."
    daffodils are like a yellow tree
13. E: -> ((#3a points at’en’ on screen)) en
    a (COM)
In the text, David has written the phrase *en gult träd* (*a yellow tree*) with a common gender indefinite article. By stating the correct form *ett gult träd* with a neuter indefinite article (line 5), Erik publicly notices a problem in the synthetic animation. Neither of the students acted on this issue when the word was written and voiced individually before the extract started, but as the voicing of the whole sentence publicly exposes the words in combination (line 3), Erik now notices the grammatical incongruence between them (line 5). The position immediately after the synthetic voicing is the first possible place in the interaction where the students can deal with and make public their evaluation of the sentence as a whole, and here, Erik acts on this possibility.

Although Erik proffers the correct form, Erik’s response to the voicing sequence in line 5 is analysed as an other-initiation of repair, since the effectuation of the repair is performed on the keyboard (line 16–17), which is under David’s control. The repair is initiated by a spoken turn by the student sitting next to the writer, which demonstrates that voicing makes the writing process publicly available to individuals other than the writer and offers interactional opportunities for intervening in each other’s writing tasks. The poem assignment is individual, so Erik’s primary task is to write his own text, but he makes use of the public voicing as a resource to intervene in, and expand David’s writing sequence.

However, David may have spotted the incongruence independently from Erik, as is demonstrated by his tapping the VOICE icon immediately after the problem noun phrase (line 4), before Erik’s noticing. His move activates voicing of the poem text from the beginning (lines 6–12). Sequentially, both students use the text animation to check whether there is a problem, and also to check if Erik’s explicit proffer fits into the sentence. When the target sentence is voiced again (line 12), Erik points directly towards the screen (*Figure 3a*), and this time he repeats the problematic article *en* (line 13). However, as a way of claiming independent knowledge about the problem’s solution and defending his role as both author and principal of the text, David pushes away Erik’s hand (*Figure 3b*) and utters the correct article *ett*. After that, he effectuates the revision on the screen’s keyboard (lines 16–17). This revision also finishes the poem.

The students do not display disagreement about the existence of a problem with the noun phrase, or about its solution. They both claim knowledge about the correct article, which indicates that this is not a process where the software helps them to learn something new about the Swedish language. Erik displays this grammatical
knowledge when he proffers the correct form, and David claims independent epistemic access when he activates the text voicing, removing David’s hand, repeating the correct grammatical form, and effectuating the correction without delay. Both students use the sequential spaces immediately after both sentence voicings to expand the sequence with actions that identify the problem, proffer correct forms, display independent knowledge, and assert epistemic rights to physically effectuate the repair on-screen.

Extract (4) below demonstrates another sequence where the students use the synthetic voicing of a written sentence as a resource for repair, but where they do not claim knowledge about the solution. The word *salt* (*salty*) is misspelled twice, but is also a problematic word choice for an orange – *söt* (*sweet*) appear as a more
relevant word – which may contribute to the outcome of the sequence. After the sentence voicing (line 1), both students expand the voicing sequence with repetitions of the pronunciation (lines 2 and 4). Both repetitions are produced with stylised prosody that marks the synthetic voicing of the word as deviant and possibly problematic.

**Extract 4.** Assignment: Collaborative text about fruits: here oranges. Letter voicing (LV),

1. SV: "smaken e lite surt och saltid"
   the taste is a litte sour and saltyed
2. M: @salt\textunderscore vid\textbar, 
   saltised
3. M: \textbackslash BACK\textbackslash STOP BACK\textbar d BACK\textbar i
4. L: @salt\textunderscore id\textbar
   saltized
5. M: \textbackslash i g NUM STOP ((letter voicings are inaudible))
6. SV: "smaken e lite surt och saltig"
   the taste is a litte sour and salted
7. (1.0)
8. M: ABC ((shifts to the letter menu))
9. M: -> \[(de e ju )
   (that’s)
10. L: -> \langle a\colon men *skit i ba\textunderscore ra.>
   yeah but don’t bother
11. *((#4 rejecting hand gesture))
12. M: \textbackslash RETURN
13. M: *((scrolls in the text, 3.0))
14. L: *((looks around the classroom))
15-20. *((six rows omitted where LA asks for the teacher, but is informed that she has left the classroom))
21. L: okej. ((LA and M looks at the assignment))
   okay

Synthetic voicing (SV). Martin (M) dictates and Laila (L) writes. Camera operator (CO).

After Martin’s stylised repetition, he immediately deletes the full stop and the last two letters in the word (line 3). After Laila’s stylised repetition (line 4), Martin adds the letters ‘i’ and ‘g’, as well as a full stop (line 5), which activates the sentence once again (line 6). After a moment of reflection, both students notice something problematic in the voicing, and they do it in overlap – Martin’s noticing is difficult to hear but Laila says *yeah but don’t bother* accompanied by a rejecting hand gesture (lines 10–11, Figure 4). These simultaneous noticing may be a response to the (misspelled) word *saltid* (the correct spelling is just *salt*, without an ending), but the students choose to move on with the next sentence rather than to identify the trouble source. Laila’s suggestion *don’t bother* asserts her role as the principal of the text, and her rights to decide on task progressivity. Martin stops editing (he breaks a new line and scrolls up and down the text), and Laila orients away from the screen (lines 13–14). At this point Laila asks for the teacher, perhaps to recruit
help (lines 15–20 omitted from transcript). When informed that the teacher has left the classroom, both students look at the assignment on the white board for the topic of the next sentence.

Extracts (3) and (4) demonstrate that the students use the synthetic voicing as an audible diagnostic tool to collaboratively identify a grammatical error or misspelling. They use the interactional space that occurs immediately after the two voicing sequences to notice something problematic, hence displaying linguistic competence. But they also demonstrate their knowledge of how to use the tool when refraining from pursuing the correction of a word which form (and perhaps meaning) they are unsure of (salt in extract 4). When faced with a second deviant pronunciation of the word, they both stop their efforts to correct the sentence and move on to the next topic in the writing assignment, showing that the progression of the writing task has precedence over achieving normatively correct form and meaning. Successful repair in the immediate sequential space after a voicing sequence seems to occur only when the students claim knowledge about the trouble’s solution (as in extract 3). When not knowing how to effectuate repair, however, the practice in extract 4 is to move on to the next sentence rather than continue making repairs on the basis of voicings alone.

3.4. Summary of findings

The analysis demonstrates how the synthetic voicings are routinely used as resources for the organisation of the students’ collaborative writing process. The students mutually monitor the voicing of a finished word or sentence and wait for its completion before moving on to the next writing unit. The students use the sequential space immediately following the voicing for various public noticings that are related to their writing, such as confirming, assessing, explaining, or correcting, and for achieving intersubjectivity that contributes to the ongoing construction of the writing assignment as a shared project. In this process, the students treat the synthetic voicings as animations of their own writings. Synthetic voicings are used as audible diagnostic tools to check if the written words and sentences are correct enough before immediately moving on, or to publicly confirm joint writing achievements, to publicly account for the contents of a sentence, or to achieve successful repair of troubles they know how to deal with. The synthetic voicings are used
as publicly shared evidential support for the correctness of written units of which the students claim to recognise the standard pronunciation. Within the interactional space that appears after synthetic voicings, the students work to enact and defend the different writing roles they were assigned in the writing task. This is done by way of multimodal resources, such as proffering correct forms, removing the hands from the screen, turning towards the other, or pushing away hands that come too close to the screen, whereby the students negotiate epistemic rights to assess the correctness of animated units, to physically alter the text, and to decide on when to recruit help or move on to the next writing units.

4. Discussion

4.1. A public resource for achieving writing as collaboration

The analysis demonstrates that students orient to the synthetic voicings as animations (Goffman 1981) of the students’ own writing and that the students use the animations as resources for achieving progression within the writing task, for achieving intersubjectivity and for noticing aspects of content and form in the written units. The synthetic animation makes certain aspects of the writing process available as a public resource for the immediate social surroundings and offers sequential opportunities for students to engage in social interaction that constructs the writing as a collaborative activity on a systematic basis. In the sequential space following synthetic voicings, students can demonstrate that they are satisfied with the task and move on. In this process, they assess the pronunciation of a writing unit publicly, justify linguistic choices to each other, or negotiate correction of micro aspects (spelling or inflection) in the voiced unit.

These results are in line with the findings of Komter (2006), Mortensen (2013) and Dalby Kristiansen (2017) on writing aloud, but also differ from writing aloud in the sense that synthetic voicings are produced by a mediating device rather than spoken by an accountable writer. The students do not hold the iPad as such accountable for trouble, but synthetic voicings are sometimes ridiculed by students when the pronunciation deviates from the expected (see Extract 4), maybe as a way of deflecting blame from the author and principal, who is one of the students present. Another difference is that synthetic voicings voice all written units systematically, and always after the actual typing is done, creating a systematic slot for post-expansion assessment. Writing aloud, in comparison, is separated from the actual writing and may be done before, during or after the typing (Gardner and Levy 2010). Writing aloud is less systematic in terms of basic organisation, as the writer may choose, for various purposes, to voice a written unit or not, and may also coordinate the speech with other on-screen activities (Gardner and Levy 2010), which create different interactional affordances in comparison with systematic synthetic voicing.

4.2. Public claims of knowing

In the interactive space that follows immediately after a synthetic voicing, students systematically face the local task of displaying an epistemic stance (Heritage 2012a) towards the voicing (and the animated written unit) in relation to the emergent writing project – the rights to publicly assess if the voicing is correct or not and to deal with the consequences of
this knowledge stance. In this position, claims of knowledge (epistemic access) are made on the basis of what the students can hear and of their own previous linguistic knowledge (epistemic status, Heritage 2012a). The voicings as such do not include diagnostic marking of spoken forms that deviate from a normative model (cf. red underlining). Therefore, the success of synthetic voicings as a support resource for the students’ writing process depends on students’ own abilities to assess the animated forms and establish this assessment in the interaction. The actions of moving on without delay (as in extract 1), confirming the voicings with response particles, or initiating repair (as in extracts 2–4) are treated as public claims of knowing that the voicing (and the animated written unit) is correct or incorrect. Compared to visual diagnostic tools such as red underlining (Čekaitė 2009; Musk 2016), audible diagnostic tools add a systematic epistemic decision-making routine to the emergent writing process, where all responses to the voicing are also acts of collaboratively performed epistemic stances towards the voiced unit, which enables treatment of that particular stance within the public arena of social interaction.

4.3. Writing roles and epistemic rights in collaborative writing

The relevant production roles in our data may be described with Goffman’s framework (1981) in terms of principal, author, and animator, referring to a variety of written and spoken production formats. These roles offer the students different epistemic rights and obligations to contribute to the writing in terms of form and content. The students mainly orient to the synthetic voicings as representations of their own writing, orienting to them as animations and attributing the role of animator to the device. The role as writer or typist (author) is achieved by being in physical control of the iPad keyboard (also see Aarsand and Melander 2016; Melander Bowden 2019). This control is demonstrated during the routine writing process but may be negotiated when students engage in effectuating repair. In the joint fruit assignment, the typist invites the dictating student (principal) to decide on form and content, thereby yielding primary access to a correct voicing to the other party. The dictating student initiates accounting for content choices, thereby claiming primary access to that domain of knowledge. In the individual Easter poem assignment, the typist is both author and principal. The classmate asserts rights to initiate repair on form but is physically excluded from effectuating repair by typing. By keeping physical control of the on-screen keyboard when responding to the synthetic voicing, the typist defends the roles of author and principal and thereby claims the right to effectuate revisions of the emergent text. In these data, the students’ display of epistemic stances, their claims of knowing or not knowing how to assess a synthetic voicing (accept, account, or repair), is collaboratively achieved by both linguistic and embodied means (Goodwin and Goodwin 2004; Goodwin 2013, 2018) and is a central driving force in enacting the different writing roles the students are assigned within the writing task.

4.4. Pedagogical implications

The analyses show that the students monitor the synthetic voicings of their writing for linguistic aspects that they can recognise and that they successfully correct problems regarding linguistic aspects that they already master. In these cases, the synthetic voicing seems to be a well-functioning audible diagnostic tool. When encountering and
identifying problems which they do not know the solution to, however, they need additional support, e.g. from the software’s spellchecker or from other persons in the classroom. On some occasions, the students try to use the synthetic voicing as a diagnostic tool in extended repair sequences (not demonstrated in the extracts above), but none of these efforts are successful in terms of resolving the problem. We plan to investigate such efforts further in a separate study.

The data demonstrate that the students trust the synthetic voicings to be accurate and authoritative representations of their writing that provide them with an audible diagnostic tool to assess their own writing process. But since this trust is based on their own abilities to recognise voiced word forms and the grammatical correctness of voiced sentences, students also need to be taught the limits of the tool. Students should become aware of the constraints of the software and adopt a critical awareness when dealing with its resources (see Čekaitė 2009; Geer et al. 2017, for similar points). This strongly suggests the relevance of developing pedagogies that enable teachers to scaffold students’ digital writing regarding both the possibilities and limits of technological tools. These results also raise some questions on the usability of synthetic voicings for newly arrived students of Swedish as a second language, or for persons with problems of language understanding and remembering (e.g. language impairment or aphasia), whose linguistic competences are limited or disabled.

**Disclosure statement**

No financial interests or benefits have arisen from the direct applications of this research.

**Funding**

This work was supported by the Marcus and Amalia Wallenberg Foundation [MAW 2014.0057].

**Notes on contributors**

**Niklas Norén** is Docent in Education at Uppsala University. Using the practices and principles of ethnomethodological and multimodal interaction analysis, his research concerns interactional aspects of developing reading and writing practices in school, the organisation of interaction involving children with communicative disabilities in everyday and institutional contexts, and young student’s digital literacy practices.

**Helen Melander Bowden** is an Associate Professor of Education at Uppsala University. Using ethnomethodology and multimodal interaction analysis, her main research interests concern knowledge and learning in interaction and the role of epistemics and affect in the unfolding organisation of action. Her work covers various areas such as learning in peer groups, instructional work in educational contexts, and interaction in professional contexts.

**Ann-Carita Evaldsson** is a Professor of Education at Uppsala University. Her research combines ethnographic studies with interactional approaches to children’s everyday lives, their peer language practices and socialisation across culturally and linguistically diverse settings. Her research includes studies of identities-in-interaction, moral work and affect, multilingualism, and games as situated activities.
References


**Appendix. Transcription conventions**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>turn</td>
<td>Natural speech.</td>
</tr>
<tr>
<td>“turn”</td>
<td>Synthetic voicing (a turn produced with text-to-speech).</td>
</tr>
<tr>
<td>🎧</td>
<td>Speaker icon marks lines with synthetic voicing.</td>
</tr>
<tr>
<td>SYMBOL</td>
<td>Writing letter by letter on the iPad’s keyboard.</td>
</tr>
<tr>
<td>SPACE=space bar, VOICE=voice icon, STOP=full stop, NUM=numbers menu, ABC=letters menu.</td>
<td></td>
</tr>
<tr>
<td>(NEUTR)</td>
<td>Keyboard icon marks lines where the keyboard is used.</td>
</tr>
<tr>
<td>(STOP)</td>
<td>Abbreviation of grammatical category in English glossings.</td>
</tr>
<tr>
<td>((nods))</td>
<td>Transcriber's difficulty to hear or see a communicative action.</td>
</tr>
<tr>
<td>turn</td>
<td>Transcriber’s description of events and actions.</td>
</tr>
<tr>
<td>turn?!</td>
<td>Emphasis in natural speech or TTS turns.</td>
</tr>
<tr>
<td>turn?</td>
<td>High pitch rise at the end of an intonation contour.</td>
</tr>
<tr>
<td>turn,-</td>
<td>Slight pitch rise at the end of an intonation contour (continuing).</td>
</tr>
<tr>
<td>turn,</td>
<td>Pitch drop at the end of an intonation contour (closing).</td>
</tr>
<tr>
<td>.h.turn</td>
<td>Inbreath. Talk on inbreath.</td>
</tr>
<tr>
<td>tfurn</td>
<td>Pitch peak in natural speech.</td>
</tr>
<tr>
<td>“turn”</td>
<td>Quiet natural speech.</td>
</tr>
<tr>
<td>“”turn“”</td>
<td>Whisper.</td>
</tr>
<tr>
<td>&gt;turn&lt; &lt;turn&gt;</td>
<td>Faster or slower speech rate.</td>
</tr>
<tr>
<td>+turn+</td>
<td>Increased volume on natural speech.</td>
</tr>
<tr>
<td>@turn@</td>
<td>Stylised voice quality.</td>
</tr>
<tr>
<td>()</td>
<td>Micro pause.</td>
</tr>
<tr>
<td>(1.0) (2.2)</td>
<td>Pauses in seconds and tenths of seconds.</td>
</tr>
<tr>
<td>tu::rn</td>
<td>Prolonged speech sound.</td>
</tr>
<tr>
<td>((gaze&gt;&gt;))</td>
<td>Prolonged embodied action.</td>
</tr>
<tr>
<td>tur rn, -</td>
<td>Onset and ending of overlapping talk. The left-pointing brackets may be excluded.</td>
</tr>
<tr>
<td>Ltu^[rn</td>
<td>Onset of overlapping embodied resource, e.g. when combining speech and head movements. Participants are assigned individual symbols (∇, *).</td>
</tr>
<tr>
<td>tuaVrn</td>
<td>Position of video screenshot in turns.</td>
</tr>
</tbody>
</table>