Identifying Engagement in Children's Interaction whilst Composing Digital Music at Home

Corey Ford
Queen Mary University of London
London E1 4NS, UK
c.j.ford@qmul.ac.uk

Nick Bryan-Kinns
Queen Mary University of London
London E1 4NS, UK
n.bryan-kinns@qmul.ac.uk

ABSTRACT

Identifying points of engagement from a person's interaction with computers could be used to assess their experience and to adapt user interfaces in real-time. However, it is difficult to identify points of engagement unobtrusively; HCI studies typically use retrospective protocols or rely on cumbersome sensors for real-time analysis. We present a case study on how children compose digital music at home in which we remotely identify points of engagement from patterns of interaction with a musical interface. A mixed-methods approach is contributed in which video recordings of children's interactions whilst composing are labelled for engagement and linked to i) interaction logs from the interface to identify indicators of engagement in interaction, and ii) interview data gathered using a remote video-cued recall technique to understand the experiential qualities of engaging interactions directly from users. We conclude by speculating on how the suggested indicators of engagement inform the design of adaptive music systems.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**; **Empirical studies in HCI**; • **Applied computing →** *Sound and music computing*; *Education*.

KEYWORDS

engagement, flow, creativity, creativity support tools, music, composition, adaptive systems, children, remote, online, novice

ACM Reference Format:
Corey Ford and Nick Bryan-Kinns. 2022. Identifying Engagement in Children's Interaction whilst Composing Digital Music at Home. In *Creativity and Cognition (C&C '22), June 20–23, 2022, Venice, Italy*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3527927.3532794

1 INTRODUCTION

The term engagement has been used in Human-Computer Interaction (HCI) to describe points where people are drawn in and attentive during interaction with a computer [11, 16, 51, 70]. Indeed, the state of mind when engaged with an interface is closely related to the psychological state of flow [11, 51, 70], where the balance of an individual's challenge and ability leads to a sense of control and loss of self-consciousness [20, 43]. Many HCI researchers have considered flow and engagement as essential aspects of the user experience [7, 9, 44, 51, 70]. However, identifying points of engagement is an ongoing HCI research challenge. Often, methods to measure engagement use retrospective protocols [50, 51, 53] or make use of sensor technology which is too obtrusive to be used outside of specialist settings [24, 39], undermining characteristic qualities of the creative experience such as spontaneity [11]. We suggest that points of engagement can be identified from a person's interaction with a user interface (UI) and present a method for doing so. Our method links points of engagement with interview data to provide insight into the user's subjective experience and may potentially be applicable to other open-ended activities without easily measurable metrics of success, such as task completion time (see [53]).

As a case study, we explore how children compose music using a digital musical interface. We examine music composition as it affords an open-ended and complex interaction in which engagement is an important quality [11, 44, 71]. Our aim is to identify relationships between users' patterns of interaction with a UI and points of engagement. This is achieved using a method which combines qualitative and quantitative methods. Patterns of interaction are identified from the actions of the children within our case study and supported by interviews based on a remote video-cued recall (VCR) technique, giving insight into the experiential qualities of their engagement with a musical interface. As a remote home-use study, with no in-person interaction between researchers and participants, we suggest that there is potential to apply our method to a range of hard-to-reach user communities. Our findings could possibly inform the design of systems which automatically adapt during interaction to support specific user's engagement. To summarise, we offer the following contributions:

- A method for identifying points of engagement using data collected solely from a UI which could be used to describe patterns of interaction for an observed set of users. This is presented as a set of metrics for identifying engagement from a case study of children interacting with a digital musical interface.
- An approach for linking interaction data and interview data via a VCR technique to gather information from users on what they did and why, reducing the need for researchers to interpret interactions based on only their intuitions (as in existing methods, see Section 2.3). This is applied in a remote study setting, potentially providing opportunities to reach many different users.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
C&C '22, June 20–23, 2022, Venice, Italy
© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9327-0/22/06…$15.00
https://doi.org/10.1145/3527927.3532794

2 RELATED WORK

Engagement is a complex, multifaceted, construct. In HCI, engagement has been defined as points where people are drawn in and attentive during interaction with a computer [11, 16, 51, 70] and has received a large amount of attention in HCI sub-fields such as funology [7] and gamification [22]. Consequently, researchers have established key attributes of engagement that can be measured and used in evaluations [11, 58, 71]. For example, O'Brien and Toms [51] proposed attributes of engagement based on extensive user studies and an interdisciplinary literature review, including: attention, novelty, control, positive affect (pleasure) and feedback.

The theory of flow [20] is frequently cited by HCI researchers when exploring attributes of engagement [1, 7, 11, 44, 47, 51]. The theory posits that people experience an optimal *flow state* characterised by nine conditions, which include balancing task challenge with skill, presenting clear goals, and providing immediate feedback [20]. With respect to composing music digitally, researchers within both HCI [6, 7, 48] and digital music education [1, 44, 47] have suggested using flow as basis for evaluating UIs designed to support enjoyment, creativity and learning. A key difference between engagement and flow is that flow must be intrinsically motivated and requires long-term focus, whereas engagement can be extrinsically motivated and sometimes occurs in the midst of multitasking [51, 70].

Some attributes of engagement, coinciding with attributes of flow, have been successfully identified via observation as opposed to through an individual's subjective self-reporting [1, 11, 21, 58], which is typically used to identify engagement. Evaluating engagement via observation can be useful in creative contexts where retrospective protocols do not necessarily capture characteristic qualities of the experience such as spontaneity [11]. In music education, Custodero [21] developed a set of observable attributes of flow states which occurred during children's daily musical experiences. Custodero's attributes are similar to O'Brien and Tom's [51] and have been used to study engagement in children's musical improvisations with a human-AI piano system [1].

Based on the described work above, notably [1, 21, 51], we explore the following attributes of engagement in our case study:

- **E1: Focused attention.** The child is attentive to a particular (musical) idea or UI element.
- **E2: Clear goals.** The child shows the intention to perform specific musical ideas such as to create an ascending motif.
- **E3: Clear cut feedback.** The child reacts to feedback from the system.
- **E4: Pleasure.** The child is enjoying themselves.

It is important to note that our intention is not to measure points of flow per say, as our interest is in the objective identification of points of engagement rather than an introspective assessment of participant's feelings that characterises the flow state [20, 43] — especially given the theory of flow's roots in positive psychology [43, 60]. We also do not focus on examining different levels of engagement (such as in [14, 51]) as our interest is in identifying where points of engagement might occur in a user's interaction with a UI, not to what extent.

2.1 Creativity Support Tools

HCI designers strive to provide opportunities for engagement with computers in many ways, such as through gamification [22]. We focus on Creativity Support Tools (CSTs) whose aims include fostering users' engagement and skill development in creative contexts [27, 63]. A set of design principles for developing CSTs from the seminal National Science Foundation workshop [63] include to enable collaboration, provide a low threshold of entry (to engage novices), and high ceiling (for expert users) [32]. The principles have been influential, inspiring the design of many CSTs [27], including in educational contexts [25, 40, 45]. However, they mainly focus on professional creativity, supporting engagement to motivate professional users in developing expertise. In contrast, Casual Creators are a sub-genre of CSTs which emphasise a non-professional's initial short term enjoyment when interacting with CSTs [18, 19]. A set of pragmatic design patterns for designing Casual Creators are offered in [19] with key patterns being to provide instant feedback and entertaining evaluations.

2.2 Music Composition & Technology

In terms of music and creativity, there is a vast amount of previous work in education and psychology examining how people learn musical constructs (such as pitch [36, 38]) or engage with music in everyday life [30, 37]. Indeed, early work on the psychology of creativity informed models of the composition process [64, 69, 73]. Wallas's [68] seminal model described the creative processes as a set of linear stages, including preparation (consciously collecting ideas and planning), illumination (the culmination of unconscious thoughts), and verification (conscious testing of an outcome). Models of the music composition processes extended this [64, 69, 73], incorporating iterative movements between phases and placing emphasis on preparation.

Numerous musical CSTs have been designed and evaluated to support the composition process. For example, Nash [44, 46] explored learning and motivation in users of a soundtracker (a grid-based music sequencer) plugin by examining 1000+ user interactions. Bryan-Kinns with various co-authors [10, 11] identified points of mutual engagement for non-musicians collaboratively creating music remotely with a simple loop-based interface. Beyond music composition, Addessi *et al.* [1, 2] explored how young children interacted with a "virtual copy of themselves" [52] in playful musical interaction — their system allows children to play with a piano and receive generated musical responses. Nijs *et al.* [47] developed a system which also extends a musical instrument, using flow and theories of embodied cognition, where performers could create artwork using their acoustic instrument and body movements.

Based on the models of the composition process and the user studies of CSTs above, we briefly outline some typical patterns of musical interaction that could be expected in children's music composition with CSTs:

- **B1: Preparation.** Children often perform a prepatory phase where they set up a scaffold for their music before starting to compose [44, 68, 69].
- **B2: Auditioning.** Children with more musical expertise with a CST likely spend a shorter amount of time editing their

music between episodes of playback [44, 46], performing edits during, and reacting to, auditory feedback [1, 19, 44, 46].

- **B3: Contemplation.** Long episodes of listening possibly indicate contemplative flow states [10] and can be described as a distinct phase [73].
- **B4: Envisioning.** Children likely envision musical ideas before committing them to notation [64, 68, 68, 69].

2.3 Approaches to Identifying Patterns of Interaction

There are many HCI studies on creativity and engagement beyond music which aim to identify patterns in people's interactions. Some rely on tasks with a clear completion goal [9, 23, 49, 53–55, 72] – for example, Pastushenko and colleagues [53] speculate on how a gamified system could detect flow states, later presenting a pilot investigation for predicting attributes of flow [49]. Other investigations have identified patterns of interaction in open-ended interfaces [35, 62, 65, 67] similar to the system used in our case study (see Section 3.1). These patterns of interaction are often determined by visualising common interactions which researchers interpret as points of engagement based on their intuitions. For example, Soltanheis [65] developed a visualisation system for interaction logs collected from an open-ended application for children's literacy and labelled examples of interactions where users acted intentionally. Our method incorporates VCR (see Section 3.4.2) to directly understand how children perceive their own interactions, which might help researchers understand how they relate to points of engagement. Furthermore, a number of these works [49, 53, 65] focus on identifying patterns to train classifiers which predict engagement based on entire sessions of interaction – we strive to identify patterns of interaction which could potentially be used to detect points of engagement whilst children are composing.

3 METHOD

We conducted a fully remote — meaning online with no in-person interaction between researchers and participants — case study of children's engagement when composing music for the first time at home with a musical CST, named *Codetta* (see Section 3.1). The purpose of the study was to explore the research question of which patterns of interaction with a musical UI might indicate points of engagement for the children in our case study. The study was designed to be fully remote to i) follow social distancing restrictions in place at the time, and ii) potentially accommodate otherwise hard-to-reach participants. The study was approved by Queen Mary University of London's ethics committee, following their standard procedures which can be found online. Both parents and children provided written consent using information sheets and consent forms designed for their reading level. For convenience, we use the term parents to refer to parents and guardians.

3.1 Codetta

Codetta [25, 26] is an open-source block-based CST for music composition. We chose Codetta in this study because: i) it is sufficiently developed and simple enough to use, meaning that children can successfully write short musical compositions [25, 26]; ii) the software contains many characteristics of CSTs such as undo features and context help documentation, supporting CST design principles [63]; iii) when talking to teachers in our recruitment pool it was apparent that block-based programming was commonly taught in the children's schools, thus they were likely familiar with Codetta's interaction style; and iv) we wish to contribute to the canon of research on how children can make music with block-based systems [25, 26, 56, 61].

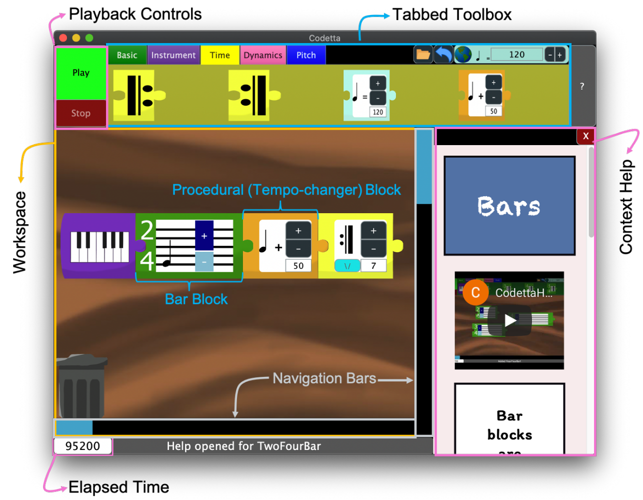

Figure 1: Screenshot of Codetta showing the composition the children create during the tutorial. The context help panel is where the tutorial is also displayed.

Codetta, labelled in Figure 1, offers a variety of blocks, collected in a tabbed toolbox. Users drag blocks from the toolbox into the workspace, where they can then combine, delete, and edit each block. Codetta offers two main types of blocks: bar blocks and procedural blocks. Bar blocks contain a stave of fixed length where children can incrementally add notes from left to right. Once added, these notes can be shifted up and down the C major scale using two arrows that appear when the mouse is hovered over. No sharp or flat notes are offered to increase the likelihood that the children's music is harmonious [26]. When pressing the play button in the top left, the bar blocks are played back sounding the connected instrument (e.g. the piano in Figure 1). Procedural blocks change a high-level aspect of the music and can create variations over time; Codetta includes procedural blocks for looping and varying the music's dynamics, tempo and pitch (see [25, 26]).

3.2 Participants

Children were recruited by approaching parents via e-mail, starting from the first author's existing contacts with schools and clubs. The e-mail contained a link to a booking page where parents could select a time that would not disrupt their everyday activities. 10 children participated in total (4 Females, 5 Males and 1 Non-binary). 6 children were 9 years old, 2 were aged 10, and 2 were aged 11. No financial incentive was offered to take part in the study.

To understand the children's background in both music and computing, we gathered self-report measures of their confidence, shown in Figure 2. The statements were adapted from [25] and

	Not confident	A little confident	Somewhat confident	Mostly confident	Super confident	I don't know
Q1: I am confident in music lessons.	O	O	O	O	O	O
Q2: I am confident in reading music notation.	O	O	O	O	O	O
Q3: I am confident with block-based computer programs (like Scratch or Purple Mash).	O	O	O	O	O	O
Q4: I am confident using a computer.	O	O	O	O	O	O
Q5: I am confident with music software (like Garage Band or Logic Pro).	O	O	O	O	O	O

Figure 2: Questionnaire used to gather participant's confidence.

extended based on preliminary discussions with the recruitment pool of teachers regarding software used in the children's schools (e.g. Q3 and Q5 reference existing software). This was selected in place of other measures such as [42] which would have been too complicated for children of this age [59]. We use a center value that is not neutral as is recommended for surveys designed for children [59], and interpret "I don't know" as no response.

The children's self-reported confidence is shown in Table 1 alongside descriptive statistics. We interpret these to mean that the children are confident with block-based tools, reasonably confident in music lessons, but not so confident with music notation or music software.

3.3 Procedure

A parent and their child joined a Zoom[1] video-conferencing call with the first author of this paper (RA) and an assistant (RB). RA is a male PhD researcher with over 5 years experience studying both music technology and composition, whilst RB is a female masters student who – although a non-musician – was completing a related project on motivation and music notations. Participants were asked to share their screen. RA then recorded the Zoom call meaning that everyone's faces and the participant's screen were captured. It was important to record the children's faces as the researchers analysed their facial expressions to identify points of engagement both during and after the study session (see Figure 3).

Codetta was sent to the participants via the video-call chat in a compressed ZIP file. The children then completed the pre-questionnaire (see Section 3.2) and followed a tutorial adapted from [25] which was extended to show the children how to use the newly developed help feature (see the tutorial in the Appendix).

Parents were asked to "busy themselves" so as to not distract the children, as suggested in [1]. The children were then given 20 minutes to compose a piece of music but were told they could stop when they wanted. The duration was based on the median time children spent composing in [25] and is similar to timings used in similar studies such as [11]. The task was purposely open-ended as we are interested in understanding how children use Codetta unprompted. Verbal help from the researchers was only given if requested and referring to something that cannot be discovered through Codetta's help documentation (for example, how to scroll the workspace).

After composing, the children were given a 10 minute break. On return, a remote video-cued recall (VCR) technique was used to interview the children about their creative process (see Section 3.4.2), lasting 10 minutes. The children were also asked to answer two post-task questions (see Section 3.4.3). The maximum study duration was 50 minutes (including break and setup time).

3.4 Data Collection

A mixed-methods approach was used to gather data capturing both 'when' and 'why' children were engaged, as described below.

3.4.1 Interaction Data. Logs of each click and drag interaction performed were generated by Codetta. Each interaction was represented by i) the elapsed time since program start in milliseconds and ii) a description of the event (for example, "piano block added" or "note moved up"). The elapsed time is truncated to the nearest tenth of a millisecond so that it was easier to jot down during the study session (see Section 3.4.2). An example log is visualised in stage 3 of Figure 3.

3.4.2 Remote Video-Cued Recall. During the study session, RA and RB used the elapsed time displayed in the bottom left of Codetta (see Figure 1) which updates on each mouse click, to identify points of engagement whilst the children were composing. To share the workload, RA observed 'E1: Focused attention' and 'E2: Clear goals', whilst RB observed 'E3: Clear cut feedback' and 'E4: Pleasure' (see Figure 3, stage 1). These labels were not used to determine our final labels – labelling of *all* attributes of engagement were also performed by both researchers independently after the study session (see Figure 3, stage 3). The time-stamps noted in this step were only used to navigate the video recordings of the children composing, which were replayed to participants during the interview by RA, remotely sharing the recorded video via their computer screen (see Figure 3, stage 2). At each time-stamp, we asked the children to comment on "what they were doing" and "why". As parents were also present, they were free to comment at any time.

3.4.3 Post-Task Questionnaire. As engagement is characterised as points where people are attentive, a very short questionnaire was presented to the children (to keep the total study time at an appropriate length [59]), capturing possible confounding variables: Q6 - "I felt under extra pressure because I was being recorded"; and Q7 - "I felt under extra pressure because my parents were in the room". The children were asked to select from an ordinal 5 point scale with the following anchors (left to right): "Not true", "I don't think so", "Maybe", "I think so" and "Very true". We reiterated to the children that we would like their honest opinion and reiterated to parents to "busy themselves" to partly mitigate the risk of misleading results.

[1] https://zoom.us/

Table 1: A breakdown of the participant's background, alongside descriptive statistics and observations based on discussions with participants during the study.

ID	Age	Gender	Q1	Q2	Q3	Q4	Q5	Observations
P1	9	Female	2	2	4	4	1	
P2	10	Female	4	1	5	4	4	Plays the piano.
P3	9	Male	5	5	4	5	1	Used music terminology with ease.
P4	9	Non-Binary	2	2	2	4	1	
P5	9	Female	4	4	4	5	3	Plays the piano. Used music terminology with ease.
P6	11	Male	3	4	5	5	.	
P7	11	Male	4	2	5	4	3	
P8	10	Female	4	4	5	5	4	Plays both trumpet and piano. Has sat graded exams.
P9	9	Male	2	1	2	5	1	Friends with P10.
P10	9	Male	2	1	4	5	4	Friends with P9.
		Mean:	3.20	2.60	4.00	4.60	2.44	
		Std-dev:	1.14	1.51	1.15	0.52	1.42	
		Median:	3.50	2.00	4.00	5.00	3.00	

1 = "Not Confident"; 2 = "Alittle Confident"; 3 = "Somewhat Confident"; 4="Mostly Confident"; 5="Super Confident"

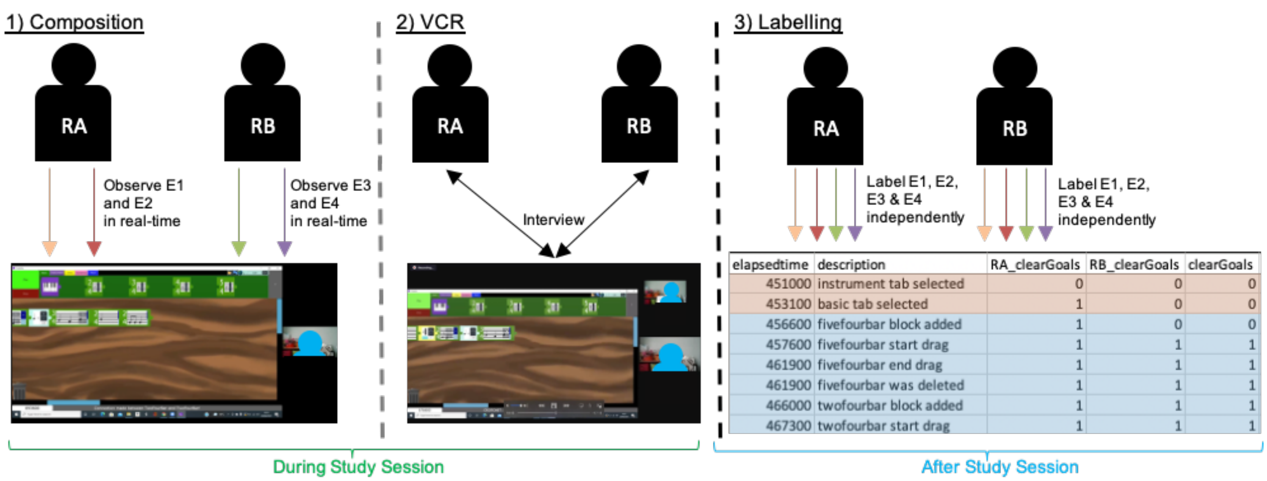

Figure 3: A visualisation of the data collection procedure. Stage 1 indicates which attributes of engagement RA and RB observed in real-time. Stage 2 indicates the VCR process (see Section 3.4.2). Stage 3 shows which attributes of engagement RA and RB labelled after the study session – a partial visualisation of a coded interaction log is also shown, highlighting interactions using the colours shown in Figure 4.

3.4.4 Labelling Attributes of Engagement. To label the interaction logs with points of engagement, we used the attributes of engagement listed in Table 2. After each study session, RA and RB reviewed the video recording of the child composing, noting start and end points for each attribute. The logs were then labelled so that each interaction performed during each attribute of engagement was marked with a 1, otherwise 0. An example section of a labelled interaction log is presented in step 3 of Figure 3 (for 'E2: Clear Goals'). RA and RB iteratively performed this procedure independently until a Cohen's Kappa comparison of these labels, for all the collected interactions, showed a reasonable degree of fit (κ = .761). We used Cohen's Kappa as it is suitable for assessing the agreement between two raters when using binary values [57]. The final points of engagement were derived by labelling where RA and RB fully agreed (i.e. where their labels overlapped for each observed attribute of engagement). Table 2 lists observations of the children which inductively emerged from this analysis. The labelled log files are in the Appendix.

3.5 Data Analysis

We used five data analysis techniques: categorising, windowing, linear mixed effect models, decision trees, and thematic analysis, discussed below.

Table 2: Examples of observations for each attribute of engagement (see Section 2).

Attribute of Engagement	Examples for Codetta
E1: Focused Attention	Child looks back and forth trying to spot a specific note during playback; Child licks lips (concentrating); Child looks for a specific block in the workspace; Child leans in to inspect their music more closley.
E2: Clear Goals	Child changes a parameter or note based on what they heard during playback; Child decides to drag in a block, decides it was the wrong move, and then reconnects another; Child clearly creates a specific shape with the notes.
E3: Clear cut Feedback	Any positive or negative reactions (e.g. "that sounds weird"); Child dances or bobs their head to the music; Child hums a tune.
E4: Pleasure	Child smiles; Child laughs; Child exclaims that something is "fun" or "cool".

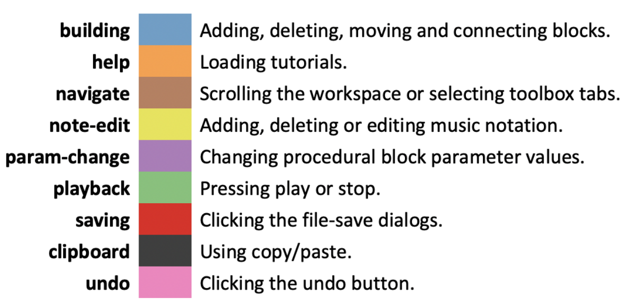

Figure 4: Coding Scheme for the interaction logs, adapted from [25].

3.5.1 Categorising & Windowing. As a first step in analysing the interaction logs, we programatically assigned each interaction to a category, inspired by [25, 44, 71]. Specifically, we used the categories developed in [25], shown in Figure 4, which were tailored to Codetta whilst being representative of other interactions common across CSTs. For the full list of categorised interactions, see the Appendix.

To uncover patterns of interaction that could be used to identify points of engagement during the composition process, we used a windowing technique, counting the percentage of interactions and engagement labels across windows of the interaction logs. A window and hop size of 25 seconds was used based on the average length of all engagement labels (22.5 seconds), which we rounded to an integer value to make calculations easier. This resulted in 348 observations which we will refer to herein as 'dataset 1' and can be found in the Appendix. We shuffle and split the data into training and validation sets to report the generalisability for models, as described below.

3.5.2 Linear Mixed Effects Models. By following the windowing technique described in Section 3.5.1, we measure the percentage of interactions performed every 25 seconds for each participant. As we are interested in which interactions with Codetta's UI might identify points of engagement, and our samples are dependent repeated measures (coming from just 10 different children), we use linear mixed effect models. These models are useful as they capture individual idiosyncrasies as random effects and are appropriate when "we only have a limited number of observations" [34, pg. 267]. Indeed, linear mixed effect models have been successfully applied in other HCI studies [29, 41, 75]. We use linear mixed effect models to predict all engagement labels and each labelled attribute of engagement as a function of our interactions, considering participant ID as a random effect. We limited interactions to the following as they occurred most frequently (see Section 4.1): note-edit, navigate, playback, building and param-change. All linear models are trained on a test-validation split of 80%:20% and we report a standard measure of accuracy (the sum of squared residuals) – see [74] for detail.

3.5.3 Decision Tree. To visualise which combinations of interactions might indicate engagement, we used a Decision Tree (DT) algorithm – a supervised machine learning technique which splits a dataset multiple times based on a target variable and can be visualised as a human interpretable flowchart. Our choice was inspired the work described in Section 2.3, but particularly motivated by [53]'s suggestion that DTs could be usefully applied to predict flow states with gamified systems. Moreover, DTs are easily interpreted by humans and identify if-then rules that can likely be quickly applied to adaptive systems. We used the CART (Classification And Regression Trees) DT algorithm detailed in [28] as it is provided by the 'rpart' library[2] and thus easily reproducible. We trained our trees on a modification of dataset 1 (named 'dataset 1b'), where engagement values are set to True if above 0, otherwise False. As in Section 3.5.2, we limited interactions to the following as they occurred most frequently (see Section 4.1): note-edit, navigate, playback, building and param-change.

As we are mostly interested in identifying children's patterns of interaction, we first trained a tree on the entirety of dataset 1b, which we name DT1. We stress here that DT1 is overfit and visualises non-generalisable patterns in our collected data. Indeed, the purpose of DT1 is to reflect idiosyncrasies in participant's interactions which might be drawn out from our qualitative analysis. Furthermore, we split dataset 1b into a training and validation set (as described in Section 3.5.1) fitting a second DT to the training set to capture information which is more generalisable, which we name DT2. DT2 is pruned to reduce its complexity and prevent overfitting by programmatically determining the improvement in error when a node is split (named the complexity parameter) with the optimum prediction accuracy as calculated by the leave-one-out cross validation procedure, described as follows: i) leave out

[2]https://github.com/cran/rpart

one data point and train the model, ii) test that model against the removed data point and record the accuracy, iii) do this for all data points and average error estimates for the final accuracy (see [74] for details). Leave-one-out cross validation is useful as we have a small dataset [74]. Furthermore, we use the validation set to report standard prediction accuracy metrics for classification, described in Table 3 (see [74] for detail).

3.5.4 Thematic Analysis. To analyse our qualitative VCR data, we used Thematic Analysis (TA) [8] to identify insights into why the children performed certain interactions whilst composing. We used TA instead of other approaches such as discourse analysis [66] because researchers in music HCI have found that a good level of descriptive detail can be obtained whilst the method is still manageable by an individual researcher [12].

The VCR portion of the Zoom recording was transcribed by RA to the level of utterances (i.e. including all "Umms" and "Uhhs") and included descriptions of the replayed videos (see transcripts in the Appendix). This provided the raw text for the inductive TA, following the steps in [8]. Completed by RA, initial codes were identified in the raw text, then grouped into themes. These themes were then reviewed and defined. RA also used the constant comparison method [17] to validate their analysis, performing the process many times and retaining the most prominent codes across iterations.

4 RESULTS

Firstly, we report analyses of the interaction data using descriptive statistics, mixed regression and decision trees (DT). Then, the Thematic Analysis (TA) [8] of the remote video-cued-recall (VCR) data is reported, giving insight into children's perceptions of their points of engagement. The post-task questionnaire is also reported.

4.1 Interaction Log Analysis

The children composed for a total of 2 hours, 24 minutes and 47 seconds (2h 24m 47s). The mean time spent composing was 14m 29s *(SD = 5m 37s)*, with values ranging between 5m 23s and 21m 17s. The mean length of all engagement labels was 22.5s, with a standard deviation of 27.1s *(n=105)*.

Figure 5 shows the amount of time children displayed points of engagement as a percentage of the total amount of time they were composing. All engagement labels occurred between 11.4% and 40.8% of the time, with a mean of 26.9% *(SD = 0.1%)*. 'E2: Clear goals' were displayed most often *(M = 13.9%, SD =7.6%)*, followed by 'E1: Focused attention' *(M=8.4%, SD=4.8%)*. 'E3: Clear cut feedback' occurred the second least *(M = 3.3%, SD = 3.1%)*. 'E4: Pleasure' occurred the least *(M = 3.8%, SD = 5.9%)*. P9 is an outlier who demonstrated E4 for 20.0% of their time composing.

Figure 6 shows the types of interactions children performed for each of the labelled attributes of engagement as a percentage of the total number of logged interactions. The children mostly performed building interactions (adding, dragging or deleting blocks), equating to 35.2% of all interactions. This was followed by navigate interactions (changing toolbox tabs or sliding around the workspace area), which equals 28.7% of all interactions. Navigate interactions also accounted for a high proportion of 'E1: Focused attention' *(49.1%)* and 'E3: Clear cut feedback' *(70.1%)*, whereas building interactions accounted for most 'E2: Clear goals' *(44.5%)*. Param-changes

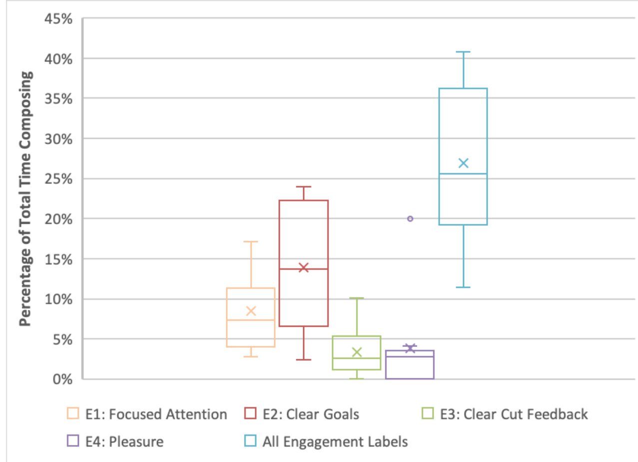

Figure 5: The proportion of time children showed points of engagement for each of the labelled attributes of engagement.

(tweaking values on procedural blocks) contributed the most to 'E4: Pleasure' *(58.7%)*. Help, clipboard and undo interactions were performed rarely *(<0.01%)*.

4.1.1 Linear Mixed Effects Models Analysis. After splitting dataset 1 into training (*n=276*) and validation (*n=72*) sets, we fit a linear mixed effects model to the training set, to predict all engagement labels and each labelled attribute of engagement, as described in Section 3.5.2. The resulting equations are:

$$Engagement = (.73)no + (.77)pa + (.58)bu + (.95)pl + (.79)na - .40$$
$$E1 = -(.14)no - (.22)pa - (.21)bu - (.10)pl - (.08)na + .26$$
$$E2 = (.50)no + (.49)pa + (.38)bu + (.35)pl + (.42)na - .25$$
$$E3 = -(.10)no - (.08)pa - (.14)bu - (.04)pl - (.03)na + .13$$
$$E4 = (.05)no + (.20)pa + (.10)bu + (.42)pl + (.10)na - .09$$

, where no = note-edits, pa = param-changes, bu = building, pl = playback and na = navigate interactions.

Playback accounts for most of the prediction for all engagement labels *(β = .95, SE = 1.50, t = .63, p = .53)* – children pressing play or stop contributes most towards predicting engagement. Conversely, building interactions accounted for the smallest amount of all engagement labels *(β = .58, SE = 1.52, t = .38, p=.70)*. Of the engagement labels, param-change was most detrimental to the prediction for 'E1: Focused attention' *(β = -.22, SE = .98, t = .23, p = .82)* – tweaking of procedural blocks negatively predicts points of focus. 'E2: Clear goals' was accounted for mostly by note-edits *(β = .50, SE = 1.37, t = .37, p = .72)* – adding, removing or editing notes might help to predict when children follow a specific strategy. Playback contributed the least to this prediction *(β = .35, SE = 1.37, t = .37, p = .80)*. Building interactions were the most detrimental to 'E3: Clear cut feedback' *(β = -.14, SE = .55, t = -.25, p = .80)* – adding, deleting or dragging blocks contribute to points where children do not openly react. Playback contributed the most to 'E4: Pleasure' *(β = .42, SE = .66, t =.64, p = .52)*, with note-edits contributing the least *(β = .05, SE = .66, t = .07, p = .94)*.

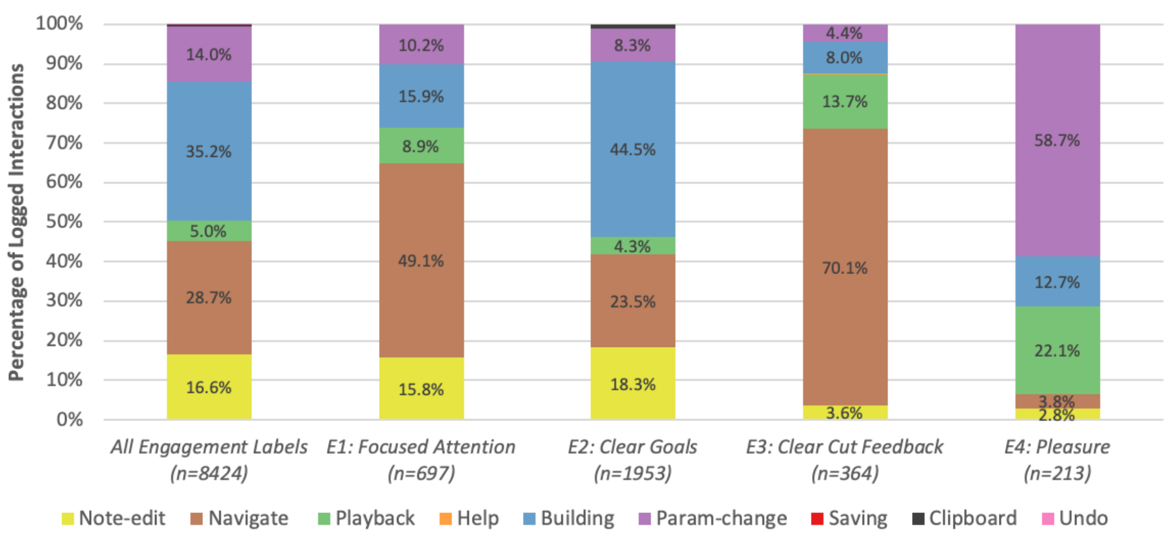

Figure 6: An overview of the children's interactions as a percentage of all logged interactions and for each labelled attribute of engagement. Note that the labelled attributes of engagement can overlap meaning that the values for each interaction across E1 through to E4 do not sum to 100%.

All of our linear mixed models performed poorly (R^2 = *0.01*, *< 0.00*, *0.06*, *< 0.00* and *0.002* for linear models of all engagement labels, E1, E2, E3 and E4 respectively). Our interactions also did not achieve statistical significance. Therefore, although the findings reported above do not represent generalisations, they might be useful in quantifying how different interactions contributed to labelled aspects of engagement for the children in our study.

4.1.2 Decision Tree Analysis. The output of DT1 is shown in Figure 7 with the leaf nodes denoted by characters ***a*** through ***d***. We also show a breakdown of each labelled attribute of engagement in each leaf node as a percentage of the total number of engagement labels, in Figure 8. Playback is the first interaction to split the tree into leaf ***a***. Leaf ***b*** is reached when the child does not perform many playback or building interactions, but is navigating the workspace — 'E2: Clear goals' and 'E1: Focused attention' account for most of these interactions (both *57.1%*). At leaf ***c***, the children do not use playback or building, but instead perform a large number of note-edits — a majority of these interactions belong to 'E2: Clear goals' (*88.9%*). Leaf ***d*** occurs when children performed a combination of building and note-edit interactions but did not use playback — these interactions mostly predicted 'E2: Clear goals' (*61.1%*).

To determine which interactions might be somewhat generalisable, the data was split into a training set and validation set, containing *n=274* and *n=74* data points respectively, with the training set being used to fit DT2. Following the procedure described in Section 3.5.3, DT2 predicts engagement based on two interactions, visualised in Figure 9: i) if playback occurred then engagement is predicted, and ii) if close to zero building interactions occurred then engagement is predicted. Table 3 reports metrics for DT2's performance calculated using the validation set. We infer that DT2 performs reasonably well, with an especially good recall.

4.2 Thematic Analysis

Fifty seven codes were identified from the TA [8] of the VCR transcripts, leading to 6 broad themes: T1 – Clear Ideas (11 codes); T2 – Ideas in Head (5 codes); T3 – Accidents (5 codes); T4 – Exploring Blocks (12 codes); T5 - Self-Assessing Music (18 codes) and T6 - Space (6 codes). We describe these below.

4.2.1 T1: Clear Ideas. The children had clear ideas about what they wanted to include in their composition [P2 - P10]. In some cases, this was described using specific elements of Codetta's notation. For example, P5 "wanted [their] piece to be in the treble clef not the bass clef". P3 explicitly noted that they "wanted[...] like[...] a four four bar", as well as "two notes and then a second note that's two beats long". They also, when shown themselves filling a bar with quaver notes, said: "I was just trying to get all the notes ready for my scale to go upwards" – implying their intention to use an ascending scale.

Other children had more general ideas about how their piece should sound. Generally speaking, the children wanted their music to be more "complex" (P6), "interesting" (P9) or "longer and[...] better" (P10). For example, P4 and P7 both wanted to include more notes within their bars, leading to compositions with a greater rhythmic density.

4.2.2 T2: Ideas in Head. Some of the children reported having ideas in their head for their music [P7 - P9]. When asked why they were using a separate block to vary note pitches, P7 said: "I just use the blocks to like [...] turn out what I thought it [in] my head was going to be the music". The idea that children used Codetta to match sounds in their head was further described by P7 when shown themselves nodding and humming along to the music: "I thought of it in my head and[...] bobbed on to what I would hopefully think it was."; "I thought of the song in my head and I was humming along

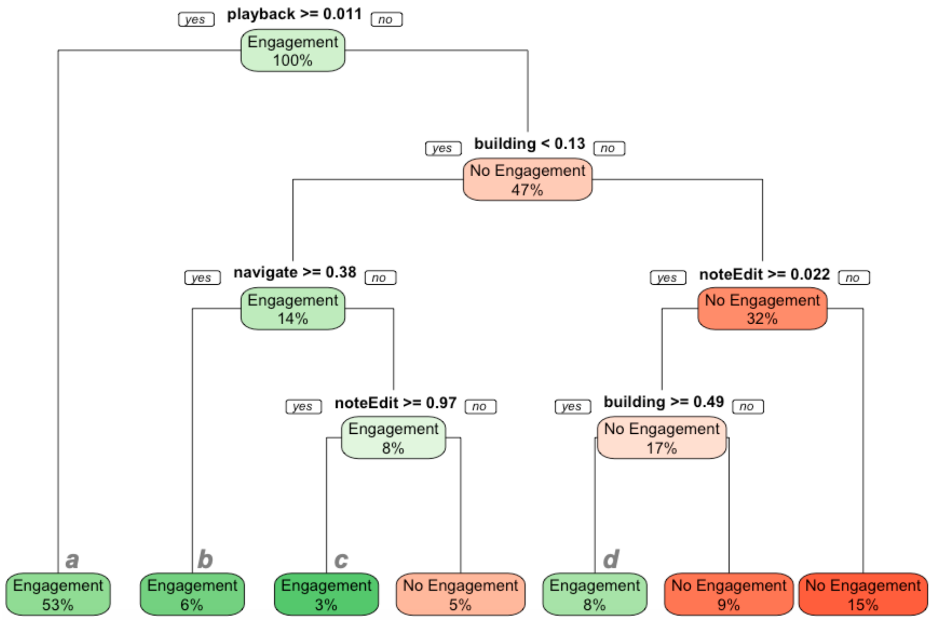

Figure 7: A decision tree fit to our entire dataset, named DT1, predicting which interactions with Codetta's UI indicate points of engagement every 25 seconds. Percentages in each leaf represent the proportion of observations in the node.

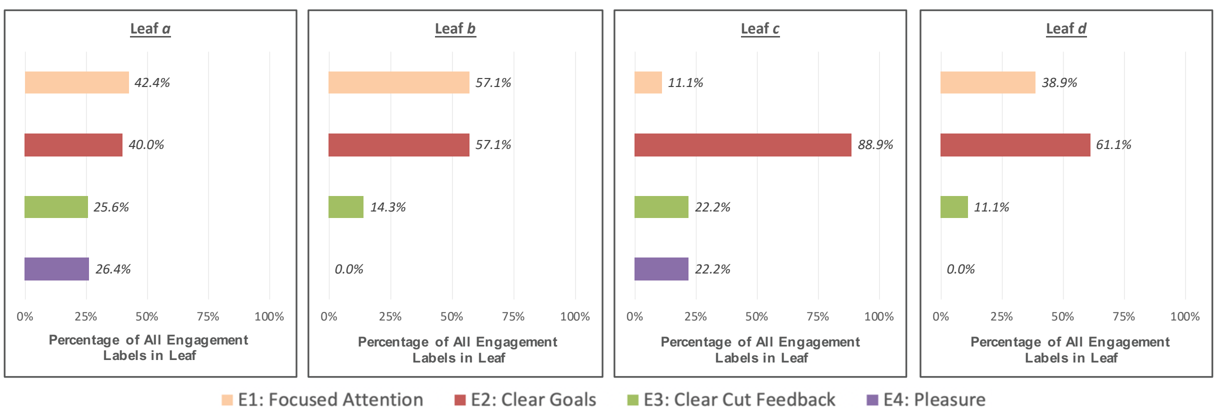

Figure 8: The amount of each labelled attribute of engagement as a percentage of the total number of engagement labels within each leaf node *a* through *d* of DT1 (see Figure 7).

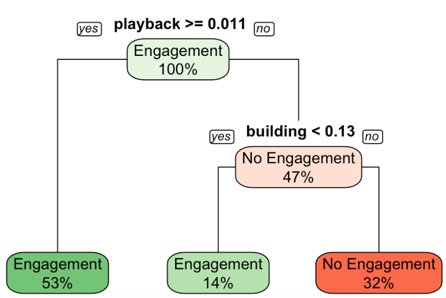

Figure 9: A pruned decision tree fit to our training data, named DT2, predicting which interactions with Codetta's UI indicate points of engagement every 25 seconds, validated in Table 3. Percentages in each leaf represent the proportion of observations in the node.

Table 3: Performance metrics for DT2, visualised in Figure 9.

Leave-One-Out DT Metrics		
Metric	**Description**	**Score**
Accuracy	Ratio of correctly predicted observations to all observations	0.77
Precision	Ratio of correctly predicted engagement observations to all correctly predicted observations	0.77
Recall	Ratio of correctly predicted engagement observations to the all engagement observations	0.85
F1	Weighted average of precision and recall	0.80

to the actual song to see if they were like connect and right with each other". When asked to comment on one of their melodies, P9 said: "I wanted it to go (singing) boom ding boom ding" and had "just made [the music] up".

P8 also developed ideas in their head, however, was less prescriptive. Following their parent's question of "did you know in your head what you wanted it to sound like?", they said that they "kind of improvised [but...] had a vague thing of what [they] wanted it to be like in [their] head".

4.2.3 *T3: Accidents.*

The children reported that they often performed interactions by accident [P1, P2, P4, P6, P8]. Commenting on themselves adjusting a note during playback before eventually deleting its bar, P1 said they "accidentally pressed on[...] the wrong note". P6 likewise "had messed up one of the notes[... so] tried to[...] get rid of that note that [they] put wrong". When discussing adding blocks from the toolbox, P2 suggested that "if [they] dragged [a bar] like to the other side that was probably an accident".

P8 adapted their piece in response to an accident. After accidentally deleting a line of music, they created an accompaniment part which was "kind of like this beat only less frequent". Commenting on this, P8 said: "I don't mind[...] because I managed to make it better by changing it".

4.2.4 *T4: Exploring Blocks.*

Many of the children tried to understand how Codetta's blocks worked by testing them [P1-P5, P7]. P2 and P4 both reported tinkering with the blocks to create some initial sounds. Furthermore, P4 "tried doing some music without [the tempo-changer block] and then with it" to figure out what the block did. P1 similarly notes that they were "just trying to see what the umm[...] blue[...] whatever it does", when dragging in the tempo-setter block.

The tempo blocks in particular seemed problematic for the children to self-assess. When testing the block, P4 said: "well that doesn't really seem to make much of a difference". P1, similarly, noted that the blue tempo-setter block "sounded the same as the orange [tempo-changer] one". P3 and P5, on the other hand, understood what the tempo blocks did, but had difficulty setting their values. P3 explicitly said they were "a bit confused what number [to] pause at for when it should be the right speed".

P7 was the only participant who successfully used the context help feature; P9 used the feature after prompting from their parent but did not follow the written instructions. When asked questions on why they sought out help, P7 stated: "I didn't know what[...] that block would do so I used the instructions". Although we observed that P7 successfully used the in-built documentation to correct their program's order of execution, they stated that "they'd learnt that the block changed like the different tone of the music".

4.2.5 *T5: Self-Assessing Music.*

The children reported checking the quality of their music [P1-P5, P7-P10]. For example, P5, on finishing their piece, decided to "check[...] for any errors". P3 also "check[ed] that the next note was [correct]". P8 similarly spent their time "trying to get [their notes] right". P4, P7, P9 and P10 all commented on alternating between high and low pitches until deciding the final pitch of a note. P9, for example, commenting on moving notes, said that they were "just doing it and then listening to it to see if it was good and if it wasn't [they'd] move it down one". P10 similarly noted that they were "going in this pattern of high low high low" and even implied that the sound triggered on moving a note provided enough information such that the play button was redundant: "when you like put them high and put them low you can kind of, you can hear the sounds, so I was basically using it off of that". P5 was the only other participant who decided not to use the playback button, saying: "I was quite worried that something would go really wrong [...] I thought just looking at it would be a better way to do it".

4.2.6 *T6: Space.*

Five children discussed how they organised and used Codetta's workspace [P1, P3, P5, P6, P8]. P5 for example "didn't notice that you could slide along so[...] thought[...] it all had to squeeze into a gap" — a comment echoed by P8. P5 thought that scrolling around "was quite easy and quite fun", whereas P1 felt that they *had* to use the navigation bars which were "quite fast".

P5 also placed all the blocks in a row, as opposed to on top of one another, because they assumed "that the parts probably wouldn't play together so[...] was trying to attach them so that they were gonna play together". In contrast, P6 purposely aligned their bars "cause, when you press play they both simultaneously start playing and[...] that would be a lot more easy to understand".

4.3 Post-Task Questionnaire Analysis

Most of the children said "Not true" *(4/10)*, "I don't think so" *(4/10)* or "Maybe" *(1/10)* in response to feeling under pressure as they were being recorded *(M=1.5, SD=1.269, n=10)*. Likewise, most children said "Not true" *(8/10)* or "I don't think so" *(1/10)* in response to feeling under pressure as their parents were in the room *(M=2, SD=1.247, n=10)*. In contrast, P9 answered both statements as "Very true".

5 DISCUSSION

We suggest that the decision trees (DT) in Figure 7 and Figure 9 might indicate patterns in the observed children's interaction with Codetta that have a relationship to their points of engagement. Measured every 25 seconds as a percentage of all performed interactions, our proposed indicators of engagement are:

- **M1: Playback**
  Children used the playback controls the majority of the time during points of engagement (> 1%) – see Figure 7 and 9.
- **M2: Building**
  Children spent less than 13% of the time adding, deleting or dragging blocks during points of engagement – see Figure 7 and 9.
- **M3: Note-edits**
  Children edited notes for more than 98% of their interactions during points of engagement – see Figure 7.

We discuss each indicator of engagement considering their relationship to themes identified in our TA [8] of interview data (see Section 4.2) and previous research, including the patterns of musical interaction listed in Section 2.2 (referred to as B1 to B4 throughout) and attributes of engagement listed in Section 2 (referred to as E1 to E4 throughout). We then reflect on our method and discuss future work.

M1: Playback

We found that playback interactions are the first interaction to divide the DTs. Indeed, playback was important in DT2 (see Figure 9) which had a reasonably good accuracy (see Table 3). T5 (Self-assessing music) suggests that the children were using playback to self-assess the quality of their musical compositions, iteratively tweaking notes until it sounded "correct" (P8). The mixed regression equations for 'E4: Pleasure' and 'E3: Clear cut feedback' also suggest that pressing play contributed towards the children reacting openly to their music. This supports B2 (Auditioning) and B3 (Contemplation), which suggest that people listen to their music to inform their decisions and make edits.

The cases of P10 and P5 in T5 (Self-assessing music) contradict the finding that playback indicates engagement because both participants did not use the play button. P10 suggested that the sounds triggered when moving a note are more helpful than the play button, whereas P5 was "worried that something would go really wrong". It is possible that playing back smaller parts of the music frequently is more important to engagement than listening to the piece as a whole, as supported by B2 (Auditioning) and the Casual Creator pattern to provide immediate feedback [19].

M2: Building

Based on the assumption that the children were largely focused and attentive when engaged (see E1 and E2 in Figure 5), we offer a possible explanation for the limited (and often negative) contribution building interactions made to the mixed linear models for engagement and 'E2: Clear goals', as well as in DT2 (see Figure 9). It is possible that dragging, adding and deleting blocks were preparatory steps that children needed to take before actively working on their music. This is supported by B1 (Preparation). However, T3 (Accidents) offers another explanation: the children would often drag in blocks by accident, likely leading to annoyance. The potential for frustration is supported also when considering T4's (Exploring blocks) finding that, even once dragged in, various blocks were difficult to understand. Coupled with the knowledge that the children are unlikely to make use of help documentation, this supports Ford *et al.*'s [25] suggestion that children must sometimes figure out how to use Codetta's blocks as they are not immediately intuitive. Nonetheless, we suggest the contemplative interactions we observed for 'E1: Focused attention', which are similar to B3 (Contemplation), sometimes indicated moments where children were reflecting on their interactions during the composition process and could present an interesting avenue for further study.

M3: Note-edits

Leaf ***c*** in DT1, where children are performing a large number of note-edits, contains mostly 'E2: Clear goals'. As note-edits were also prominently observed as 'E2: Clear goals' during labelling (see Table 2), we suggest it is possible that the children needed to manipulate the music at note-level to realise their musical ideas. Indeed, we suggest the finding that children spent most of their time performing 'E2: Clear goals' (see Figure 5) is a notable interaction for our sample. T1 (Clear Ideas) supports this, showing that children have specific ideas about what they want to include in their compositions. This notion supports the CST design principle to provide a high ceiling (affording fine grained, note-level, control) [63], but contradicts the Casual Creator principle to limit user control [19]. It is important to note that the predictive power for M3 is less so than the previous suggested metrics, and is not represented in our most generalisable model, DT2. Nevertheless, based on B4 (Envisioning) and our discussion related to reflection above, their might be a back and forth relationship between points of engagement, where the observed children use note-edits to notate ideas and validate them with playback, which is potentially worthy of further investigation.

5.1 Reflections on Method & Limitations

Due to the length of time needed to complete the qualitative portion of our analysis, we could only realistically recruit a small number of participants, limiting the statistical power of our quantitative analysis. Our use of decision trees and mixed linear models helped identify engaging patterns of interaction within our sample of children, but we cannot be confident of their generalisability. Furthermore, it may be that other models such as Support Vector Machines or Naive Bayes classifiers, as used in [9, 65], could perform well with our interaction data – we open-source our dataset to encourage further experimentation (see Appendix). Nonetheless, we found that the quantitative data analysis approach was useful for us in complementing our thematic analysis, possibly offering insights for designers and evaluators of a range of similar systems. Unlike related approaches (see Section 2.3), these insights are gleaned directly from our study population, reducing the need for researchers to solely interpret patterns of interaction. Indeed, we suggest that others could use this method to develop their own metrics for their target users or system — the thematic analysis reveals "rich and detailed" [8, pg. 78] information on the children's perceptions of their points of engagement, with the quantitative analysis complementing this data.

The nature of child-computer interaction and engagement makes it difficult to account for confounding variables. For example, although our questionnaire responses indicate that parents had a small influence, this possibly introduced distractions, making it harder for the children to engage fully. Or, as we observed in the case of P9, who was distracted by their parent, this led to outlier activity. Future work could incorporate our measures of confounding variables as random effects into multilevel models similar to our use of mixed regression. Specifically, Bayesian multilevel models [13] might be able to adequately model such effects (see [5]).

Lastly, as the researcher's backgrounds are crucial in influencing the assignment of engagement labels, a comparison between the labels and children's post-hoc responses for a questionnaire on engagement (e.g. [33]) would have improved the study's external validity. This was omitted from this study due to time restrictions. Labelling could be performed by professional educators who are close to the children's world, ensuring social context is captured.

5.2 Future Work

In terms of designing an adaptive CST for children, it might be possible to use the suggested indicators of engagement to automatically adjust a UI and possibly better support engagement. We suggest design ideas for adapting Codetta's interface below based on our identified metrics, being careful to take into account our small sample size and that the way in which Codetta might adjust must be sensitive to the affective and motivational context, not causing frustrations detrimental to engagement [44].

For M1 (Playback), it may be beneficial to increase children's awareness of how often they press play. A subtle approach could be to detect when a child has performed a low number of playback interactions and use this as a trigger for gradually making the play button more translucent, based on similar HCI studies that have used fading techniques to draw attention to UI features [3, 4]. Alternatively, we could perhaps start playback automatically to prompt children — although this could be too obtrusive and detrimental to engagement. We could also take inspiration from Casual Creators [19] and provide *entertaining evaluations* to provide motivation as children self-assess their work (cf. T4 and T5). For example, a fun animation could reward children for using playback.

Advances in generative music [15, 31] also have the potential to contribute in the music composition process, supporting engagement through human-AI collaboration and co-creation. Perhaps, Codetta could interject with a block containing novel music if M2 (Building) is met, creating a comparable experience to Addessi *et al.*'s system [1, 2]. There could also be opportunities to introduce musical material that is tangential to the children's personal composition style (for example, introducing fragments of music outside of western traditions) to explore whether the children would adopt stylistic ideas similar to the AI.

6 CONCLUSION

This paper presented a fully remote method for identifying points of engagement for a group of children's interactions whilst composing with a digital musical interface at home. Through a case study of children composing music with Codetta [25] — a musical block-based CST — we successfully identified a set of indicators of engagement which, although not generalisable, describe patterns of interaction which could potentially be applied to programmatically detect their engagement whilst composing. We also successfully linked interaction data and interview data using a remote VCR technique, providing qualitative support for our engagement indicators gleaned directly from the children, lessening the need for researchers to interpret interactions based soley on their intuitions. As a remote-study, we also suggest that our method has the potential to be applied to otherwise hard-to-reach groups such as those who cannot easily attend user studies in person at a University. Overall, this work contributes towards the design of musical CSTs which might better facilitate children's learning experiences in music, and helps towards supporting HCI researchers in designing adaptive CSTs.

APPENDIX

All appendix material can be found online at: https://github.com/thecoreyford/Identifying-Engagement.

ACKNOWLEDGMENTS

Thanks to Amy Shaw for the support in data collection and labelling, Courtney Reed and Berker Banar for reading over early drafts, and the reviewers for their helpful feedback. Corey Ford is a research student at the UKRI Centre for Doctoral Training in Artificial Intelligence and Music, supported by UK Research and Innovation [grant number EP/S022694/1].

REFERENCES

[1] Anna Rita Addessi, Laura Ferrari, and Felice Carugati. 2015. The Flow Grid: A Technique for Observing and Measuring Emotional State in Children Interacting with a Flow Machine. *Journal of New Music Research* 44, 2 (April 2015), 129–144. https://doi.org/10.1080/09298215.2014.991738

[2] Anna Rita Addessi and François Pachet. 2005. Young children confronting the Continuator, an interactive reflective musical system. *Musicae Scientiae* 10, 1 (2005), 13–39. https://doi.org/10.1177/1029864906010001021

[3] Pengcheng An, Saskia Bakker, Sara Ordanovski, Ruurd Taconis, Chris L.E. Paffen, and Berry Eggen. 2019. *Unobtrusively Enhancing Reflection-in-Action of Teachers through Spatially Distributed Ambient Information.* Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300321

[4] Jeremy N. Bailenson, Nick Yee, Jim Blascovich, Andrew C. Beall, Nicole Lundblad, and Michael Jin. 2008. The Use of Immersive Virtual Reality in the Learning Sciences: Digital Transformations of Teachers, Students, and Social Context. *Journal of the Learning Sciences* 17, 1 (2008), 102–141. https://doi.org/10.1080/10508400701793141

[5] Dale J Barr, Roger Levy, Christoph Scheepers, and Harry J Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language* 68, 3 (2013), 255–278.

[6] Benjamin B. Bederson. 2004. Interfaces for staying in the flow. *Ubiquity* (September 2004). https://dl.acm.org/doi/10.1145/1029383.1074069

[7] M.A. Blythe, K. Overbeeke, A.F. Monk, and P.C. Wright. 2004. *Funology: From Usability to Enjoyment* (1 ed.). Springer, Dordrecht. https://doi.org/10.1007/1-4020-2967-5

[8] V Braun and V Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[9] Eli T Brown, Alvitta Ottley, Helen Zhao, Quan Lin, Richard Souvenir, Alex Endert, and Remco Chang. 2014. Finding Waldo: Learning about Users from their Interactions. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1663–1672. https://doi.org/10.1109/TVCG.2014.2346575

[10] Nick Bryan-Kinns and Fraser Hamilton. 2012. Identifying mutual engagement. *Behaviour & Information Technology* 31, 2 (Feb 2012), 101–125. https://doi.org/10.1080/01449290903377103

[11] Nick Bryan-Kinns, Patrick G. T. Healey, and Joe Leach. 2007. Exploring Mutual Engagement in Creative Collaborations. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition* (Washington, DC, USA) *(C&C '07)*. Association for Computing Machinery, New York, NY, USA, 223–232. https://doi.org/10.1145/1254960.1254991

[12] Nick Bryan-Kinns, Wei Wang, and Yongmeng Wu. 2018. Thematic Analysis for Sonic Interaction Design. In *Proceedings of British HCI 2018*. BCS Learning and Development Ltd., 1–3. https://doi.org/10.14236/ewic/hci2018.214

[13] Paul-Christian Bürkner. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of statistical software* 80 (2017), 1–28.

[14] Linda Candy and Zafer Bilda. 2009. Understanding and Evaluating Creativity. In *Proceedings of the Seventh ACM Conference on Creativity and Cognition* (Berkeley, California, USA) *(C&C '09)*. Association for Computing Machinery, New York, NY, USA, 497–498. https://doi.org/10.1145/1640233.1640376

[15] Filippo Carnovalini and Antonio Rodà. 2020. Computational Creativity and Music Generation Systems: An Introduction to the State of the Art. *Frontiers in Artificial Intelligence* 3 (2020). https://doi.org/10.3389/frai.2020.00014

[16] Peter McFaul Chapman. 1997. *Models of engagement: Intrinsically motivated interaction with multimedia learning software.* Ph. D. Dissertation. University of Waterloo.

[17] Kathy Charmaz. 2006. *Constructing grounded theory: A practical guide through qualitative analysis.* SAGE Publications Inc.

[18] Katherine Compton. 2019. *Casual Creators: Defining A Genre of Autotelic Creativity Support Systems.* PhD Thesis. University of California Santa Cruz. https://escholarship.org/uc/item/4kg8g9gd

[19] Kate Compton and Michael Mateas. 2015. Casual Creators. In *Proceedings of the Sixth International Conference on Computational Creativity (ICCC 2015)*, Hannu Toivonen, Simon Colton, Michael Cook, and Dan Ventura (Eds.). Brigham Young University, Park City, Utah, 228–235. http://computationalcreativity.net/iccc2015/proceedings/10_2Compton.pdf

[20] Mihály Csíkszentmihályi. 1990. *Flow: The Psychology of Optimal Experience.* Harper Collins, New York, USA.

[21] Lori A Custodero. 1998. Observing Flow in Young Children's Music Learning. *General Music Today* 12, 1 (Oct 1998), 21–27. https://doi.org/10.1177/104837139801200106

[22] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining "Gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (Tampere, Finland) *(MindTrek '11)*. Association for Computing Machinery, New York, NY, USA, 9–15. https://doi.org/10.1145/2181037.2181040

[23] Andrew Emerson, Fernando J Rodríguez, Bradford Mott, Andy Smith, Wookhee Min, Kristy Elizabeth Boyer, Cody Smith, Eric Wiebe, and James Lester. 2019. Predicting Early and Often: Predictive Student Modeling for Block-Based Programming Environments. *International Educational Data Mining Society* (2019).

[24] Cem Eteke, Hayati Havlucu, Nisa İrem Kırbaç, Mehmet Cengiz Onbaşlı, Aykut Coşkun, Terry Eskenazi, Oğuzhan Özcan, and Barış Akgün. 2018. Flow From Motion: A Deep Learning Approach. arXiv:1803.09689 [cs.LG]

[25] Corey Ford, Nick Bryan-Kinns, and Chris Nash. 2021. Creativity in Children's Digital Music Composition. In *Proceedings of New Interfaces for Musical Expression (NIME) 2021.*, Roger Dannenberg and X Xiao (Eds.). Shanghai. https://nime.pubpub.org/pub/ker5w948/

[26] Corey Ford and Chris Nash. 2020. An Iterative Design 'by proxy' Method for Developing Educational Music Interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME) 2020*, Romain Michon and Franziska Schroeder (Eds.). Birmingham City University, Birmingham, UK, 279–284. https://doi.org/10.5281/zenodo.4813361

[27] Jonas Frich, Lindsay MacDonald Vermeulen, Christian Remy, Michael Mose Biskjaer, and Peter Dalsgaard. 2019. Mapping the Landscape of Creativity Support Tools in HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, New York, USA, 1–18. https://doi.org/10.1145/3290605.3300619

[28] Jerome Friedman, Trevor Hastie, Robert Tibshirani, et al. 2001. *The elements of statistical learning* (2 ed.). Vol. 1. Springer, New York, USA. https://web.stanford.edu/~hastie/ElemStatLearn/

[29] Julian Frommel, Martin J. Dechant, and Regan L. Mandryk. 2021. The Potential of Video Game Streaming as Exposure Therapy for Social Anxiety. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 258 (oct 2021), 25 pages. https://doi.org/10.1145/3474685

[30] Alinka E. Greasley and Alexandra Lamont. 2011. Exploring engagement with music in everyday life using experience sampling methodology. *Musicae Scientiae* 15, 1 (2011), 45–71. https://doi.org/10.1177/1029864910393417

[31] Dorien Herremans, Ching-Hua Chuan, and Elaine Chew. 2017. A Functional Taxonomy of Music Generation Systems. *Comput. Surveys* 50, 5, Article 69 (Sept. 2017), 30 pages. https://doi.org/10.1145/3108242

[32] T. Hewett, M. Czerwinski, M. Terry, J. Nunamaker, L. Candy, B. Kules, and E. Sylvan. 2005. Creativity support tool evaluation methods and metrics. In *Creativity Support Tools.* 10–24. http://www.cs.umd.edu/hcil/CST/Papers/creativitysupport.pdf

[33] Susan A Jackson, Andrew J Martin, and Robert C Eklund. 2008. Long and Short Measures of Flow: The Construct Validity of the FSS-2, DFS-2, and New Brief Counterparts. *Journal of Sport and Exercise Psychology* 30, 5 (2008), 561–587. https://doi.org/10.1123/jsep.30.5.561

[34] Maurits Kaptein. 2016. Using generalized linear (mixed) models in HCI. In *Modern Statistical Methods for HCI*. Springer, 251–274.

[35] Samad Kardan. 2017. A data mining approach for adding adaptive interventions to exploratory learning environments. *University of British Columbia* (2017).

[36] Alexandra Lamont. 1998. Music, Education, and the Development of Pitch Perception: The Role of Context, Age and Musical Experience. *Psychology of Music* 26, 1 (1998), 7–25. https://doi.org/10.1177/0305735698261003

[37] Alexandra Lamont. 2011. University students' strong experiences of music: Pleasure, engagement, and meaning. *Musicae Scientiae* 15, 2 (2011), 229–249. https://doi.org/10.1177/1029864911403368

[38] Alexandra Lamont and Ian Cross. 1994. Children's Cognitive Representations of Musical Pitch. *Music Perception* 12, 1 (10 1994), 27–55. https://doi.org/10.2307/40285754

[39] Marco Maier, Daniel Elsner, Chadly Marouane, Meike Zehnle, and Christoph Fuchs. 2019. DeepFlow: Detecting Optimal User Experience From Physiological Data Using Deep Neural Networks. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, {IJCAI-19}*. International Joint Conferences on Artificial Intelligence Organization, 1415–1421. https://doi.org/10.24963/ijcai.2019/196

[40] John Maloney, Mitchel Resnick, Natalie Rusk, Brian Silverman, and Evelyn Eastmond. 2010. The scratch programming language and environment. *ACM Transactions on Computing Education* 10, 4 (2010), 16. http://doi.acm.org/10.1145/1868358.1868363.

[41] Gloria Mark, Shamsi T. Iqbal, Mary Czerwinski, and Paul Johns. 2014. Bored Mondays and Focused Afternoons: The Rhythm of Attention and Online Activity in the Workplace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3025–3034. https://doi.org/10.1145/2556288.2557204

[42] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart. 2014. The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population. *PLOS ONE* 9, 2 (02 2014), 1–23. https://doi.org/10.1371/journal.pone.0089642

[43] Jeanne Nakamura and Mihaly Csikszentmihalyi. 2009. Flow theory and research. In *Handbook of positive psychology*, Charles Richard Snyder, Shane J Lopez, Lisa M Edwards, and Susana C Marques (Eds.). Oxford University Press, 195–206.

[44] Chris Nash. 2011. *Supporting virtuosity and flow in computer music.* PhD Thesis. University of Cambridge.

[45] Chris Nash. 2014. Manhattan: End-User Programming for Music. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, K. Tahiroğlu, R. Fiebrink, A. Tana, and B Caramiaux (Eds.). Goldsmiths, London, 221–226. https://www.nime.org/proceedings/2014/nime2014_371.pdf

[46] Chris Nash and Alan Blackwell. 2012. Liveness and Flow in Notation Use. In *Proceedings of the International Conference on New Interfaces for Musical Expression 2012*, Brent Gillespie (Ed.). University of Michigan, Ann Arbor, Michigan. https://www.nime.org/proceedings/2012/nime2012_217.pdf

[47] Luc Nijs, Bart Moens, Micheline Lesaffre, and Marc Leman. 2012. The Music Paint Machine: Stimulating Self-monitoring Through the Generation of Creative Visual Output Using a Technology-enhanced Learning Tool. *Journal of New Music Research* 41, 1 (2012), 79–101. https://doi.org/10.1080/09298215.2011.650180

[48] Donald A. Norman. 1993. *Things that make us smart: defending human attributes in the age of the machine.* Addison-Wesley Pub. Co. 290 pages.

[49] Wilk Oliveira, Kamilla Tenório, Juho Hamari, Olena Pastushenko, and Seiji Isotani. 2021. Predicting students' flow experience through behavior data in gamified educational systems. *Smart Learning Environments* 8, 1 (2021), 1–18.

[50] Wilk Oliveira, Armando Toda, Paula Palomino, Luiz Rodrigues, Seiji Isotani, and Lei Shi. 2019. Towards Automatic Flow Experience Identification in Educational Systems: A Theory-Driven Approach. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (Barcelona, Spain) *(CHI PLAY '19 Extended Abstracts)*. Association for Computing Machinery, New York, NY, USA, 581–588. https://doi.org/10.1145/3341215.3356311

[51] Heather O'Brien and Elaine G. Toms. 2008. What is user engagement? A conceptual framework for defining user engagement with technology. https://doi.org/10.14288/1.0107445

[52] François Pachet. 2003. The Continuator: Musical Interaction With Style. *Journal of New Music Research* 32, 3 (2003), 333–341. https://doi.org/10.1076/jnmr.32.3.333.16861

[53] Olena Pastushenko, Wilk Oliveira, Seiji Isotani, and Tomáš Hruška. 2020. A Methodology for Multimodal Learning Analytics and Flow Experience Identification within Gamified Assignments. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '20)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3334480.3383060

[54] Thomas W Price, Yihuan Dong, and Tiffany Barnes. 2016. Generating Data-Driven Hints for Open-Ended Programming. *International Educational Data Mining Society* (2016).

[55] Thomas W. Price, Yihuan Dong, and Dragan Lipovac. 2017. ISnap: Towards Intelligent Tutoring in Novice Programming Environments. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (Seattle, Washington, USA) *(SIGCSE '17)*. Association for Computing Machinery, New York, NY, USA, 483–488. https://doi.org/10.1145/3017680.3017762

[56] Michael Quigley and William Payne. 2021. Toneblocks: Block-based musical programming. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Shanghai, China, Article 45. https://doi.org/10.21428/92fbeb44.46c0f6ef

[57] Priya Ranganathan, C S Pramesh, and Rakesh Aggarwal. 2017. Common pitfalls in statistical analysis: Measures of agreement. *Perspectives in clinical research* 8, 4 (2017), 187–191. https://doi.org/10.4103/picr.PICR_123_17

[58] Janet Read and Stuart Macfarlane. 2002. Endurability, Engagement and Expectations: Measuring Children's Fun. In *Interaction Design and Children, Shaker Publishing*. Shaker Publishing, 1–23.

[59] Janet C. Read and Stuart MacFarlane. 2006. Using the Fun Toolkit and Other Survey Methods to Gather Opinions in Child Computer Interaction. In *Proceedings of the 2006 Conference on Interaction Design and Children* (Tampere, Finland) *(IDC '06)*. Association for Computing Machinery, New York, NY, USA, 81–88. https://doi.org/10.1145/1139073.1139096

[60] Martin EP Seligman and Mihaly Csikszentmihalyi. 2014. Positive psychology: An introduction. In *Flow and the foundations of positive psychology*. Springer, 279–298.

[61] R. Benjamin Shapiro, Rebecca Fiebrink, Matthew Ahrens, and Annie Kelly. 2016. BlockyTalky: A Physical and Distributed Computer Music Toolkit for Kids. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Queensland Conservatorium Griffith University, Brisbane, Australia, 427–432. https://doi.org/10.5281/zenodo.1176120

[62] Roi Shillo, Nicholas Hoernle, and Kobi Gal. 2019. Detecting Creativity in an Open Ended Geometry Environment. *International Educational Data Mining Society* (2019).

[63] Ben Shneiderman, Gerhard Fischer, Mary Czerwinski, Mitch Resnick, Brad Myers, Linda Candy, Ernest Edmonds, Mike Eisenberg, Elisa Giaccardi, Thomas T. Hewett, Pamela Jennings, Bill Kules, Kumiyo Nakakoji, Jay Nunamaker, Randy Pausch, Ted Selker, Elisabeth Sylvan, and Michael Terry. 2006. Creativity Support Tools: Report From a U.S. National Science Foundation Sponsored Workshop. *International Journal of Human-Computer Interaction* 20, 2 (2006), 61–77. https://doi.org/10.1207/s15327590ijhc2002_1

[64] John Sloboda. 1985. *The musical mind: The cognitive psychology of music*. Oxford University Press, Oxford.

[65] Mina Soltangheis. 2017. *From children's play to intentions: a play analytics framework for constructionist learning apps*. Ph. D. Dissertation. Massachusetts Institute of Technology.

[66] Dan Stowell, Mark D. Plumbley, and Nick Bryan-Kinns. 2008. Discourse Analysis Evaluation Method for Expressive Musical Interfaces. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Genoa, Italy, 81–86. https://doi.org/10.5281/zenodo.1179631

[67] Ivan Sysoev, James H. Gray, Susan Fine, Sneha Priscilla Makini, and Deb Roy. 2022. Child-driven, machine-guided: Automatic scaffolding of constructionist-inspired early literacy play. *Computers & Education* 182 (2022), 104434. https://doi.org/10.1016/j.compedu.2022.104434

[68] G. Wallas. 1926. *The art of thought*. J. Cape, London.

[69] Peter R Webster. 2002. Creative Thinking in Music: Advancing a Model. In *Creativity and Music Education*, T. Sullivan and L. Willingham (Eds.). Canadian Music Educators Association, Edmonton, 16–33.

[70] Yongmeng Wu. 2018. *Design and Evaluate Support for Non-musicians' Creative Engagement with Musical Interfaces*. Ph. D. Dissertation. Queen Mary University of London.

[71] Yongmeng Wu and Nick Bryan-Kinns. 2017. Supporting Non-Musicians' Creative Engagement with Musical Interfaces. In *C&C '17 Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. ACM, Singapore, 275–286. https://doi.org/10.1145/3059454.3059457

[72] Kai Xu, Alvitta Ottley, Conny Walchshofer, Marc Streit, Remco Chang, and John Wenskovitch. 2020. Survey on the Analysis of User Interactions and Visualization Provenance. *Computer Graphics Forum* 39, 3 (2020), 757–783. https://doi.org/10.1111/cgf.14035 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14035

[73] Betty Anne Younker. 2000. Thought Processes and Strategies of Students Engaged in Music Composition. *Research Studies in Music Education* 14, 1 (2000), 24–39. https://doi.org/10.1177/1321103X0001400103

[74] Mohammed J Zaki and Wagner Meira Jr. 2020. *Data Mining and Machine Learning: Fundamental Concepts and Algorithms*. Cambridge University Press.

[75] Elif Ecem Özkan, Tom Gurion, Julian Hough, Patrick G.T. Healey, and Lorenzo Jamone. 2021. Specific hand motion patterns correlate to miscommunications during dyadic conversations. In *2021 IEEE International Conference on Development and Learning (ICDL)*. 1–6.