ABSTRACT

Title of Thesis:	CONTEXT AND THE CONCEPTUAL BUILDING BLOCKS FOR SYNTHESIS IN THE LITERATURE REVIEW PROCESS
	John Stroman Morabito, Master of Science, Human-Computer Interaction, 2020
Thesis Directed By:	Dr. Joel Chan, Assistant Professor, College of Information Studies

Conducting a literature review involves trying to capture knowledge in such a way that it can be reused during synthesis and writing and this can be a challenge. We set out to describe the information capture process with the goal of identifying patterns which could influence the design of software that would support the literature reviewing process. We conducted a qualitative user study of four participants with a protocol analysis of the information capture portion of their literature reviewing process, focusing on a detailed description of how they captured contextual information in their notes and annotations, and how this varied across tools used. Our analysis revealed three common patterns of context capture, and quantitative and qualitative differences in these patterns across tools for literature reviewing. These findings provide insights for system design to support information capture in literature reviewing systems.

CONTEXT AND THE CONCEPTUAL BUILDING BLOCKS FOR SYNTHESIS IN THE LITERATURE REVIEW PROCESS

by

John Stroman Morabito

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2020

Advisory Committee:
Dr. Joel Chan, Chair and Advisor
Dr. Katrina Fenlon
Dr. Wayne Lutters

Acknowledgements

Sincerest gratitude to my family, friends, and my advisor Joel, without whom none of this would have been possible.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Literature reviewing is a difficult, but essential task for scholarly knowledge work. In his paper investigating whether innovation is becoming more challenging, Ben Jones says, "...if one is to stand on the shoulders of giants, one must first climb up their backs, and the greater the body of knowledge, the harder this climb becomes" (Jones, 2009, p.284). This “climb” proves to be challenging no matter what the scale or domain. Systematic reviews, a narrowly focused, but more rigorous version of literature reviewing, are known to be updated infrequently because of how laborious they are to complete. For example, Petrosino (1999) estimates that systematic reviews can take 5-6 people more than 1000 hours to complete and this is one reason why they are rarely updated even when they need to be (Ervin 2008). Conversely, PhD students working at presumably a much smaller scope, also struggle with the literature reviewing. In an examination of expectations for doctorate level scholarship, specifically on the expectation for the literature review section, Holbrook (2008) found frequent examiner mentions of significant issues with literature review sections of even dissertations that were deemed “acceptable”. Examples of these comments include:

> “. . . the candidate takes the existing literature at face value, and rarely takes issue with a published study in terms of its methodology or interpretation. Greater critical insight would be expected at doctoral level.” (p.1032) and
>
> “There is no critical discussion of any shortcomings of the studies cited nor any critical interpretation of the significance of the findings. Similarly, there is no discussion of the

shortcomings of cross-sectional studies in general. . . . are more in the format of a 'facts presentation'." (p.1032)

The purpose of this research is to inform the design of tools that augment the literature reviewing process. Researchers use different tools, strategies, and workflows to navigate through the process. Traditionally, researchers have worked with tools that are only useful for searching for and/or collecting papers (ex. Zotero), and do not feature any extra affordances to support synthesis. In their 2019 paper, Qian mentions the lengths that researchers have to go through, adapting "literature review" tools and combining them with other tools to approximate a smoother synthesis experience (Qian et al. 2019). There is an emergence of tools that are attempting to do this, such as LiquidText (https://www.liquidtext.net/) and NVivo (https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home) (featured in this study), and we would like to know from a scientific standpoint if the affordances they provide can actually improve the literature reviewing process. To understand how we can make the process easier, we will look at literature reviewing from a conceptual perspective in addition to analyzing the workflows and tools used by researchers in situ.

One way to think about literature reviewing is as a collection of information behaviors. Information is the primary unit that researchers interact with in the literature review process, so by looking at the different ways information is handled we can begin to see patterns. Broadly speaking, the two primary behaviors in literature reviewing are information capture, where information is collected from a source, and information reuse, where an attempt is then made to combine or repurpose the collected information in a new way.

There is a significant body of research in computer-supported cooperative work (CSCW) that suggests that context plays an important role in information reuse. Generally speaking, context is the information or circumstances that surround a focal point, and it can be a common source for pain points in cases where it is missing, mishandled or mismanaged. Many of the challenges related to information reuse relate to the ability of an individual to obtain context for the information to interpret or understand it. Information may lose context over time which can make reusing it difficult. For example, in a field study of collaborative information reuse in aircraft technical support, engineers lamented reusing old records because information was missing, outdated, or not appropriate anymore because of procedural changes. Over the years if any changes to the records were not tagged, the context of those changes were lost (Lutters and Ackerman 2007). On the other hand, we see many benefits of supporting context. A study on calendar systems showed that the availability of contextual information is useful for information scanning and information retrieval (Dourish et al. 1993). In a case study comparison between the two calendar systems they note that having metadata for event information such as the title of the event or arrival time of the speaker, in addition to who the author of the information is, are critical for the interpretation of the events.

Manually adding context, to support information reuse and retrieval, can be a challenge. Anderson, Hardstone, Procter, and Williams (2008) explore this in a field study of the information management systems in the healthcare setting. Healthcare workers used a system which featured a drop-down menu that would allow them to select relevant contextual information to attach to a file where they recorded patient interactions. Often times though, it was observed that the options in the drop-down menu were not specific enough or too many options were available, which led to the employees choosing the closest option, or a catch all

"Not Specified" option. This all led to ambiguous contextual information capture which could prove problematic down the line. In another study, Hinrichs showed another situation where manual context capturing proved problematic. In both of Hinrich's field studies, observing documentation in a steel mill and in a sewerage plant, the workers faced challenges reusing documents which were outdated or inappropriate. The use of physical documents and decentralized information management meant that multiple copies of the same document could appear in different locations, each with varying levels of accuracy or completeness. In one extreme case, this "incomplete or inaccessible documentation" could lead to "costly exploratory 'digging by hand' to avoid damaging power lines" (Hinrichs, Pipek, and Wulf, 2005, p.375).

Another reason that manual context capture is hard is that the person capturing the information must predict or estimate what information will be useful for recontextualization in the future. This concept of adding contextual information to help information further down the line is explored thoroughly in Ackerman and Halverson's (2004) study of information reuse in telephone hotline groups. Proactively adding contextual information to an information object in anticipation of that contextual information's future relevance is explored with the term "trajectory." The success of recontextualization of phone records depended largely on the previous editors' ability to understand potential reuse scenarios and "make assumptions about the record's trajectory" (Ackerman and Halverson 2004 p.183).

A study by Knight, Wilson, Brailsford, and Milic-Frayling (2009) demonstrates how context can be difficult to extract when creating systematic reviews. While conducting a cognitive work analysis of medical systematic reviews they discovered that contextual information was particularly hard to extract from papers for the review, going so far as to say that systematic reviews are so painful because the reviewers are "enslaved to the trapped data",

referring to the current state of inflexibility and immobility of knowledge and information. The reviewers must collect a variety of details about each paper they review including sample sizes, demographics, experimental conditions and more, all contextual information to help interpret the claims or veracity of the paper.

In summary, in many settings where knowledge work and information reuse takes place (aircraft technical support, healthcare, steel mill, sewerage plant) there are particular pain points that revolve around the idea of context. The specific aim of this study is to understand whether and how challenges related to context manifest in literature reviewing. One study that comes close to exploring this in the academic setting looked at graduate students conducting research in libraries (O'Hara, Smith, Newman, and Sellen 1998). They do a good job of investigating the mechanics of reading and note-taking, but do not analyze them explicitly in terms of the concepts of context and information reuse. We would like to investigate in more detail the mechanics of context capture for literature reviewing, including how these mechanics interact with the tooling used for information capture. A detailed description of these mechanics could help us identify possible pain points and design ideas for how we can better support context in literature reviewing. To understand what context capture looks like in the literature review process, this study explores the following research questions through a protocol analysis of researchers conducting their literature reviews:

> **RQ1:** How are people describing context with the information that they capture and create for literature reviewing?
>
> **RQ2:** How do patterns of context capture vary across literature review tools?

Chapter 2: Methodology

To address our research questions, we conducted a detailed protocol analysis of information capture in the literature reviewing process for four researchers. Our specific focus was to describe the role of context in this process from the perspective of the tooling and behaviors of each participant.

2.1 Participants

	Education	Domain	Tools
P1	PhD	Educational Research	OneNote, NVivo, Adobe Acrobat
P2	PhD	Pain (Health)	LiquidText
P3	PhD	Social Science	OneNote, Paper, Markers, Pencil
P4	PhD	Teachers, Technology, Security, Privacy	Google Docs, Apple Built-In PDF reader

Table 1 Participant Profiles

We initially recruited 10 participants from a large public research institution in the northeastern United States through convenience selection strategies including fliers and emails (both personal, and through department channels, such as mailing lists). The participants ranged from graduate level to more senior level researchers. We narrowed down the selection of participants to four: two using tools with specialized features for literature reviewing (P1 using NVivo, P2 using LiquidText) and two using tools without special literature reviewing affordances (P3 and P4). This gave us an opportunity to go in depth into how the mechanics of context capture might vary by tool. Table 1 provides descriptive characteristics of the participants. For the rest of the study we will refer to the first group as the "specialized group" and the second group as the "generic group" in order to differentiate between the percieved difference in how much their tools are designed to support literature reviewing.

2.2 Data Collection

We invited participants to three data collection sessions (Guided Tour, Capture Session, Reuse Session). In the first session (Guided Tour), the participants gave an in-person tour to one of the members of the research team, showing the physical and digital workspace that they occupy while literature reviewing. Then in the second session (Capture Session), the participant was observed conducting the portion of their literature review where they capture information from their selected papers. A third and final session (Reuse Session) was conducted sometime after the capture session. In this session participants were asked to revisit their captured information from the previous session as they normally would. The reuse session was not analyzed in this study but is relevant to mention because it covers the expected future reuse of the captured data.

For each of the three sessions, the participants were instructed to wear a hat mounted GoPro camera to record their actions in the usual location that the participants conducted their literature reviews to maintain as close of an experience and environment to what they normally experience. For some participants this was at home and others at their office on campus. The capture sessions each lasted approximately 45 minutes. At each capture session the participant was asked to actively work on their own literature review project in an entirely self-directed manner without interference from our researcher. The goal of this was to ensure that we could specifically observe information capture at the beginning of the literature review process. One researcher administered the session, asking the participants to think aloud to verbalize their thoughts and actions during their literature review process.

2.3 Data Preparation

The footage of the information capture sessions was analyzed using qualitative coding in DataVyu, and then exported to google sheets where further analysis was conducted. Figure 1 below shows what the qualitative data coding for information capture events looks like in DataVyu (left sidebar) and the image on right is the view of the GoPro footage.

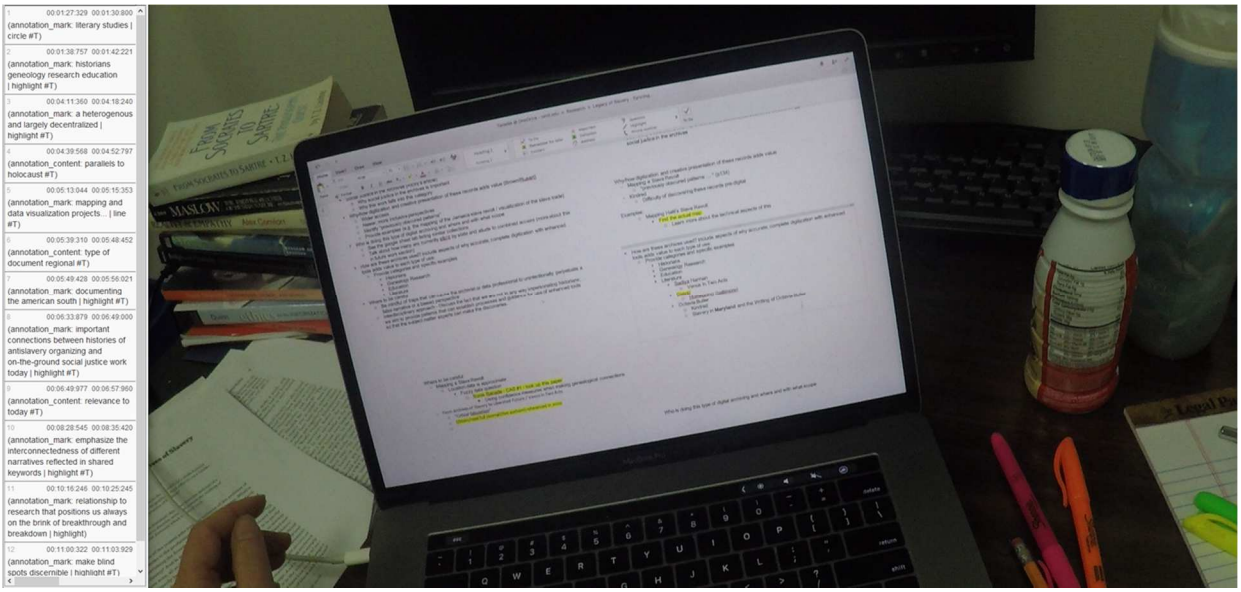


Figure 1 Coding of capture events in DataVyu (left). View of GoPro footage (right)

2.3.1 Identifying Conceptual Building Blocks

The first step of analysis was to identify when participants were capturing information. Broadly speaking, participants captured information in two forms: durable capture events, and ephemeral capture events. Durable capture events are externally observable and leave a trace or record that persists in some media such as making a highlight or writing a margin note. Ephemeral information capture events included verbalized intentions or mental notes said by the participant. An example of this would be when a participant makes a "note to self" or reminder to find a piece of information. For this study we focused on the durable information capture events because they are the most tangible events that would allow us to observe the participants

interactions with their tools. In Datavyu, we noted every durable capture event that took place during the 45 minute capture session.

After all durable capture events were catalogued in google sheets, we began to explore our research question. With our goal being to describe the mechanics of context capture in the literature review process a first necessary step was to identify what we mean by context. Context is the objects and circumstances surrounding a focal point which are used for its interpretation or reuse. Dourish gives us a good framework for thinking about context when he says, "It is not simply the case that something is or is not context; rather, it may or may not be contextually relevant to some particular activity" and "contextuality is a relational property that holds between objects or activities" (Dourish 2004). This means that we are always talking about context as something related to something else, some focal point. So we need to know what are the focal points in the literature review process for which contextual information is important.

One of the end goals of a literature review is to synthesize information. Synthesis is the combining of parts to create a new whole, such as a theory or argument. In this study we label these 'parts' as Conceptual Building Blocks. Conceptual Building Blocks (CBBs) take many forms, but all boil down to a core identifiable property: the ability to be used in a synthesis. CBBs then act as the focal points for which contextual information is important.

The main heuristic we used to identify CBBs was whether the information could be used as a building block in a synthesis. Examples of this type of information that we considered included ideas, concepts, findings, questions, theories etc. Figure 2 shows some examples of conceptual building blocks.

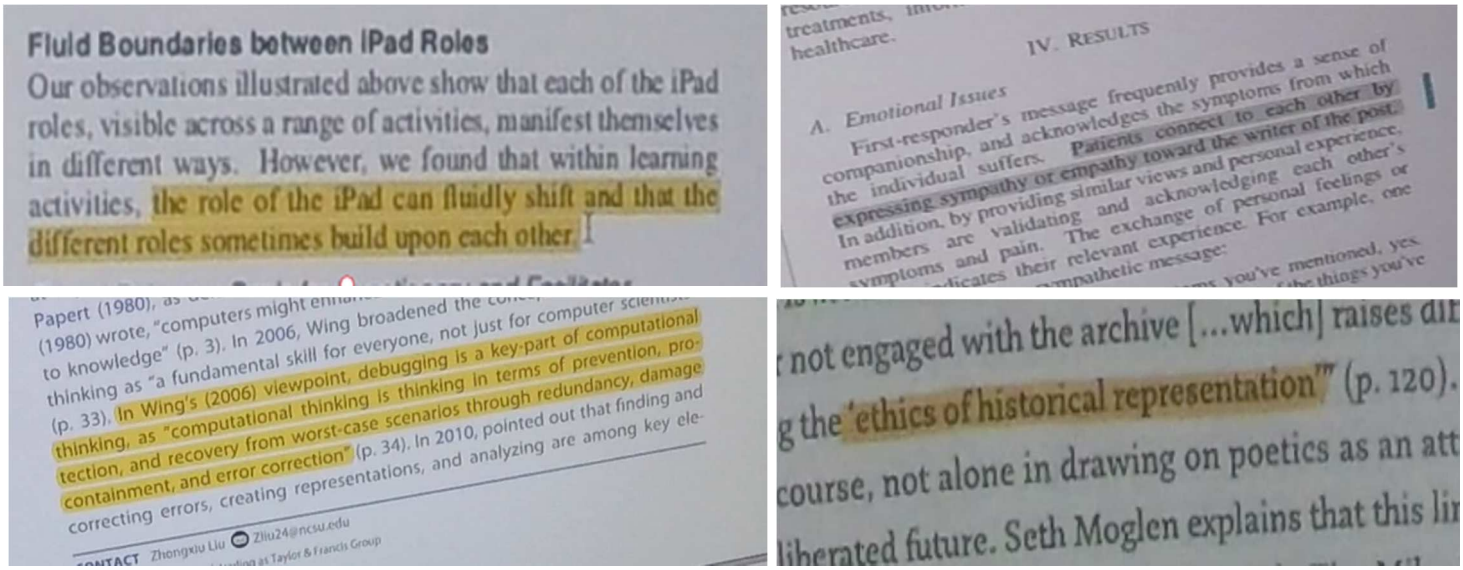


Figure 2 P4 claim highlighted in PDF reader (top left), P2 result highlighted in LiquidText (top right) P1 claim highlighted in NVivo (bottom left), P3 concept highlighted on paper (bottom right)

Beyond this, we also used cues from each participants' think aloud commentary as well as knowledge of their particular research goals, to inform which information we could label as something they would potentially use later in a synthesis. The following are examples of some of these verbal cues which hinted at information being potentially useful for synthesis:

"I feel like we may have seen this a little bit" – P4 relating an observation from their literature review to their own work.

"I have an issue here" - P1 disputing a finding.

"I've got another piece to this that I have somewhere" – P3 suggesting that this is a concept that they can add to.

2.3.2 Identifying Context Capture Events

With the conceptual building blocks identified as the informational focal points, we could then identify, relationally, what was contextually relevant to them. The following heuristics emerged as different kinds of contextual information:

1. Methods Information: this is information, typically from the methods section, that describes how the study was conducted.

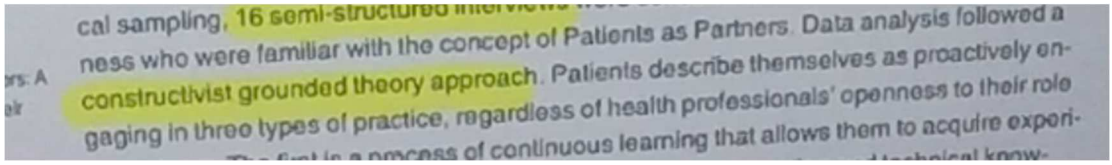


Figure 3 Two highlighted examples of methodological information from P2 using LiquidText

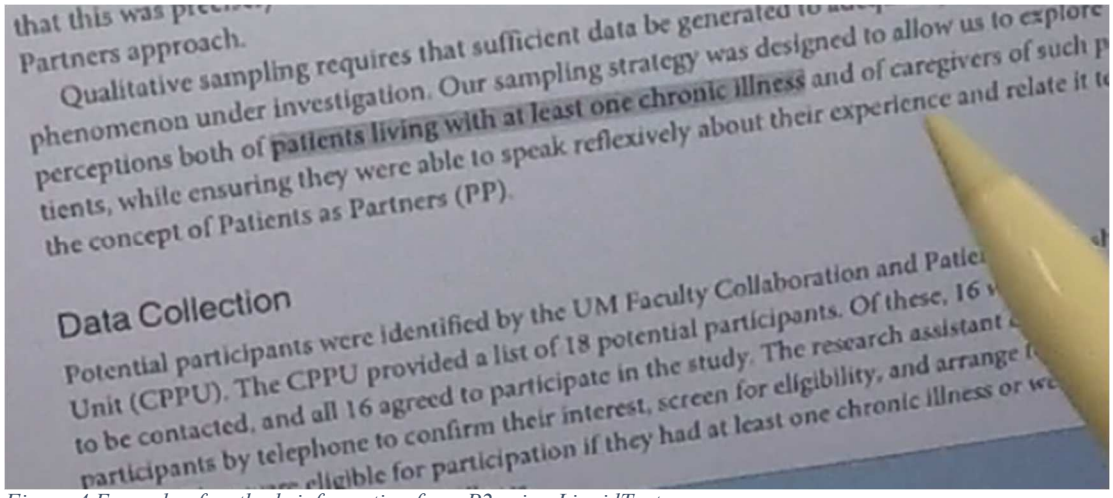


Figure 4 Example of methods information from P2 using LiquidText

2. Metadata: data that describes or gives information about other data or information. Examples observed include author name, page number, in-text citation, figure numbers and more

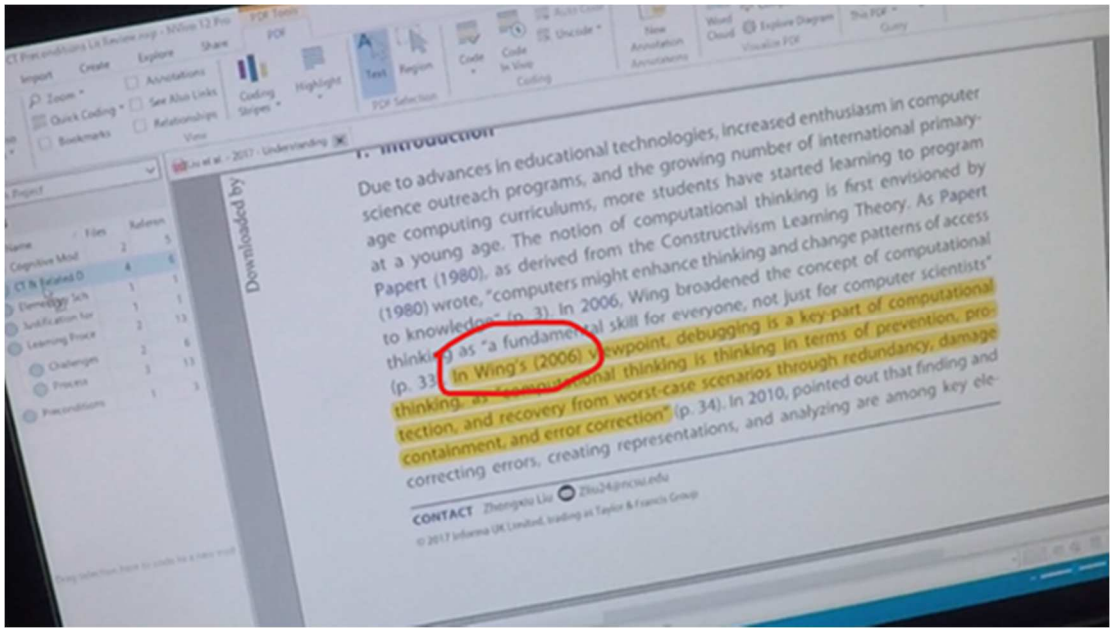


Figure 5 In-text citation inside of highlight from P1

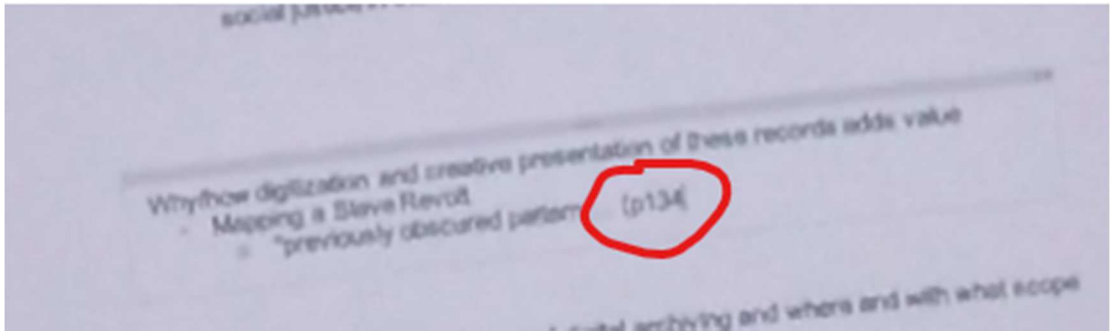


Figure 6 Page number written down with quote from paper from P3

3. Background Information: typically information that can be used to help orient the reader into the setting of the study (ex. back up or give support to claims)

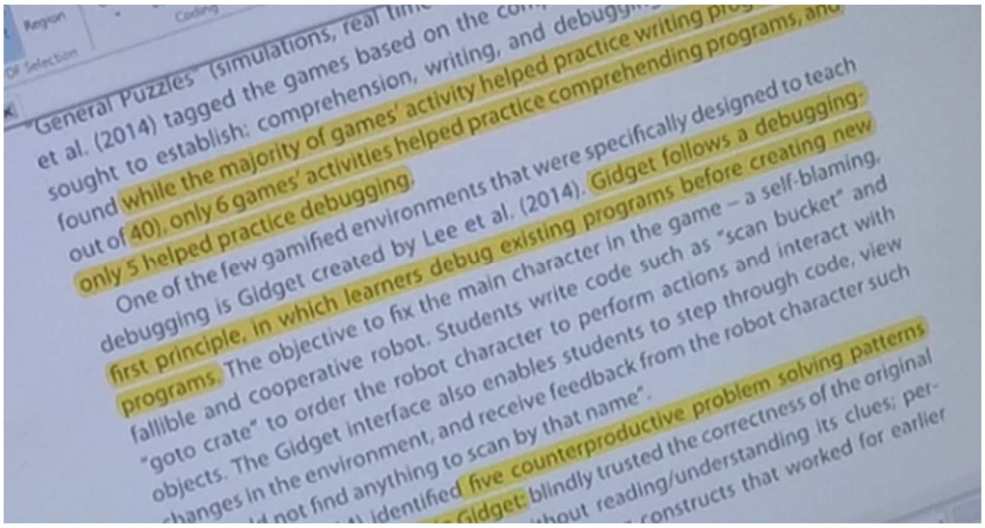


Figure 7 Middle highlight is background information from the study from P1

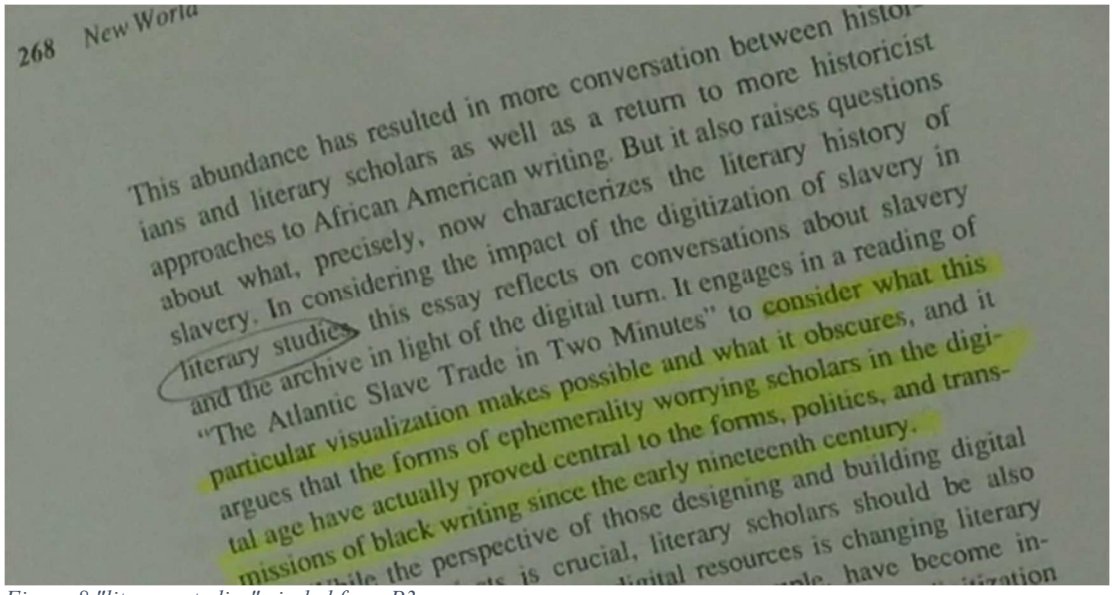


Figure 8 "literary studies" circled from P3

2.3.3 Analyzing Context Capture Events

The previous steps yielded a set of context capture events, many associated with CBBs. For the main analysis, we then iteratively examined each of these context capture events using an open coding approach, to identify recurring higher-level patterns of context capture. This part of the analysis took place in Lucidchart (see figure 9) and was organized more methodically in google

sheets (figure 10). It was in these two programs that we compared the frequency and qualitative nature of these patterns across different literature review tool groups.

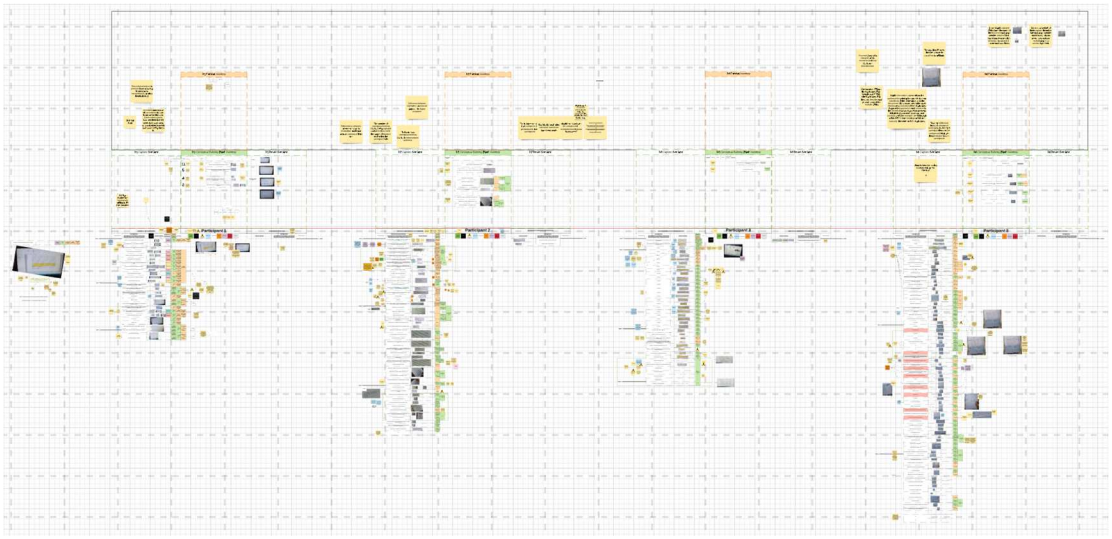


Figure 9 Overview of analysis in Lucidchart, the capture events of each participant are represented in each T shape

	A	B	C	D	E	F	G	H	J	M	P	Q	R	S	T	U	V	W	X	Y	Z
1	P_id		CBB, Cont	capture method	time coi	com	conte	context description	context captured (	proximal context	inst	notes	Conte	No Cc	Expli	Proxi	Signs	CBBs	Disembedded CBBs		Contex N
2	P1	P1_o1	CBB	NVivo Tag	10.54		in wi	in-text citation (au	in-text citation (au	page number	y	[[NVivo Tag]] sent	1	-	-	1	-	1	1		10.54
3	P1	P1_o2	CBB	NVivo Tag	6.55		shallow problem identification				y		1	-	-	-	-	1	1		6.55
4	P1	P1_o3	CBB	NVivo Tag	7.05		a deeper identification				y		1	-	-	-	-	1	1		7.05
5	P1	P1_o4	CBB	NVivo Tag	14.07		in the context of debugging				y		1	-	-	-	-	1	1		14.07
6	P1	P1_o5	CBB	NVivo Tag	5.96		selecting strategy				y		1	-	-	-	-	1	1		5.96
7	P1	P1_o6	CBB	NVivo Tag	8.88		taking more effort to understand the problem deeply				y		1	-	-	-	-	1	1		8.88
8	P1	P1_o7	CBB	NVivo Tag	5.39		locat author (proximal)			author (proximal)	y		1	-	-	1	-	1	1		5.39
9	P1	P1_o8	Context	NVivo Tag	7.29		debu background information			in-text citation, tag	y		-	-	1	-	-	-			
10	P1	P1_o9	CBB	NVivo Tag	8.87		five counterproductive problem solving pat			in-text citation, tag	y		1	-	-	1	-	1	1		8.87
11	P1	P1_o10	CBB	NVivo Tag	12.53		explore, tinker, refine			in-text citation, tag	y		1	-	-	1	-	1	1		12.53
12	P1	P1_o11	Context	highlight	6.88		the f research questions						-	-	1	-	-	-			
13	P1	P1_o12	CBB	highlight	11.61		in the gaming context, SGD provided studer			in-text citation, an	y		-	1	-	1	-	1	1		1
14	P1	P1_o13	CBB	NVivo Annotate	54.78		in the gaming context the curriculum was design for CT itself, so is that why they saw l						-	1	-	-	-	1	1		5
15	P1	P1_o14	S2C	highlight	4.71		learn methods						-	-	-	-	1	-			
16	P1	P1_o15	Context	highlight	5.16		cond methods (or maybe background infor			annotations in anr	y		-	-	1	-	-	-			
17	P1	P1_o16	CBB	NVivo Annotate	20.58		the gaming environment required some conditionals, did the robotics obstacle course						-	1	-	-	-	-			2
18	P1	P1_o17	Context	highlight	5.66		the v methods	hyperlink to appendix	section in highlight				-	-	1	-	-	-			
19	P1	P1_o18	CBB	NVivo Tag	21.59								1	-	-	-	-	1	1		21.59
20	P1	P1_o19	CBB	NVivo Tag	3.75		Appendix 2: Computational thinking Rubric			tags in tag folder			1	-	-	-	-	1	1		3.75
21	P1	P1_o20	CBB	highlight	2.71		or						-	-	-	-	-	-			
22	P1	P1_o21	CBB	highlight	2.92		and						-	-	-	-	-	-			
23	P1	P1_o22	CBB	highlight	0.87		agent and background are non-traditional a			annotations in anr	y		-	1	-	-	-	1	1		
24	P1	P1_o23	CBB	NVivo Annotate	23.27		seems to me like changing the agent and the background is not a different						-	1	-	-	-	1	1		2
25	P1	P1_o24	CBB	highlight	8.5		use of pop gaming culture			annotations in anr	y		-	1	-	-	-	1	1		
26	P1	P1_o25	CBB	NVivo Annotate	5.53		not a CT component						-	1	-	-	-	1	1		
27																					
28	P2	P2_o1	Context	OneNote note	19.22		critic article title and authors last name						-	-	-	1	-	-	-		
29	P2	P2_o2	S2C	OneNote note	24.97		use r S2C						-	-	-	-	1	-	-		
30	P2	P2_o3a1	CBB	highlight	11.39		previ no context				y	1st instance of CBE	-	1	-	-	-	1	1		1
31	P2	P2_o3a2		sticky note	5.99								-	-	-	-	-	-			
32	P2	P2_o3a3		highlight (blue)	3.21								-	-	-	-	-	-			
33	P2	P2_o3b	CBB	OneNote note	35.7		new instance, new article title and page number				y	2nd instance of CB	1	-	-	-	-	1	-		35.7
34	P2	P2_o4	CBB	OneNote note	19.69		context for o3a,b						-	1	-	-	-	1	-		1
35	P2	P2_o5	S2C	highlight (green)	7.04		S2C						-	-	-	-	1	-	-		
36	P2	P2_o6	S2C	OneNote note	27.95		S2C						-	-	-	-	1	-	-		
37	P2	P2_o7	S2C	OneNote note	6.87		S2C						-	-	-	-	1	-	-		
38	P2	P2_o8	Context	OneNote note	5.86		title of paper (met title of paper						-	-	-	1	-	-	-		
39	P2	P2_o9	CBB	OneNote note	6.66		locat title of paper (proximal context)						-	-	-	-	1	-	1	-	
40	P2	P2_o10	CBB	OneNote note	4.78		fuzzy title of paper (proximal context)						-	-	-	-	1	-	1	-	
41	P2	P2_o11a	S2C	OneNote note	38.07								-	-	-	-	-	1	-		
42	P2	P2_o11b	S2C	highlight (yellow)	8.63		"sometimes I highlight things that need action"						-	-	-	-	-	1	-		
43	P2	P2_o12	Context	OneNote note	93.27		author and metadata						-	-	-	1	-	-	-		
44	P2	P2_o13	S2C	margin note	21.59		task written in margin						-	-	-	-	-	1	-		
45	P2	P2_o14	CBB	highlight (blue)	13.74		captured citation number						1	-	-	-	-	-	-		13.74
46	P2	P2_o15a	CBB	highlight (orange)	15.48		instance						-	1	-	-	-	1	1		1
47	P2	P2_o16	CBB	OneNote note	7.68		no context						-	1	-	-	-	1	-		
48	P2	P2_o15b	CBB	OneNote note	28.18		instance	title of paper					1	-	-	-	-	1	-		28.18
49	P2	P2_o17a	S2C	OneNote note	24.42		obtai S2C						-	-	-	-	1	-	-		
50	P2	P2_o17b	S2C	highlight (yellow)	5.1		obtai S2C						-	-	-	-	1	-	-		
51	P2	P2_o18	CBB	highlight (blue)	11.46		place author highlighted in the area citation number captured						1			1					

Figure 10 Sample view of analysis in google sheets

Chapter 3: Findings

3.1 Basic Descriptives

To set the context for our findings, we first look at whether there are any context events to analyze. In total we observed 133 context related capture events. These context events were associated with a total of 108 CBBs across the participants. This means that participants engaged in substantial amounts of context capture activity that we can analyze for our research questions. In the following sections, we dive deeper into these context capture events to identify patterns of context capture as well as variations in these patterns across tools.

	CBBs	Context events
Total	108	133
Average	27	33.25
Specialized		
P1	20	37
P2	27	30
Generic		
P3	25	8
P4	36	58

Table 2 Breakdown of information capture events

3.2 Varieties of Context Capture Patterns

Context is captured in a variety of ways. Below we describe four different varieties that we observed across our four participants with examples of each.

3.2.1 Integrated Context Capture

One way that people capture context is that they integrate it into the CBB itself. In these events, context such as page number, author name, in-text citation, etc. are being captured at the same

time as a conceptual building block is captured or created. When we say context captured we mean that 1) metadata is captured and/or 2) the original context of the information object is captured by the tooling of the participant.

In P1's session we observed 11 information capture events where context was captured in addition to a conceptual building block. All 11 of these capture events are in NVivo using the tag function to capture the context of the CBB.

The following is an explanation of how NVivo captures context with its tagging system. P1 selects content and drags it into a tag folder which they name. The content that has been tagged is left highlighted blue in its original location, while a disembedded instance is sent to the tag folder. The tag folder can contain multiple tagged content excerpts but only link back to the paper they are from, not back to the specific location of excerpts.

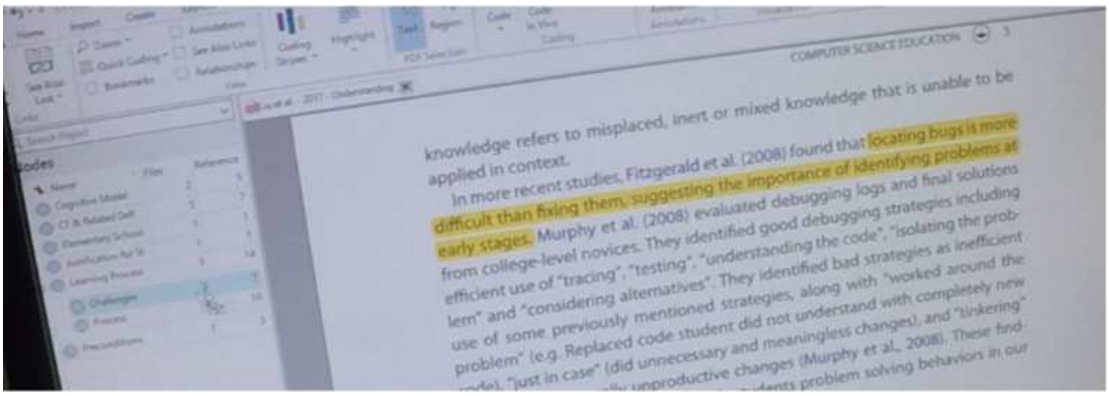


Figure 11 Example of a tag in NVivo

Here we can see in this example the highlighted area in yellow was selected then dragged and dropped in the 'Challenges' tag folder. These tag folders are created by the participant and can hold multiple tagged excerpts.

For P3 context is not captured automatically with any of their information capture methods. We observed five capture events in their capture session where they manually integrated context into the capture of a conceptual building block. For two of the five captured

CBBs P3 also captured metadata such as author name and page number. For the other three CBBs, P3 captured a citation number for each.

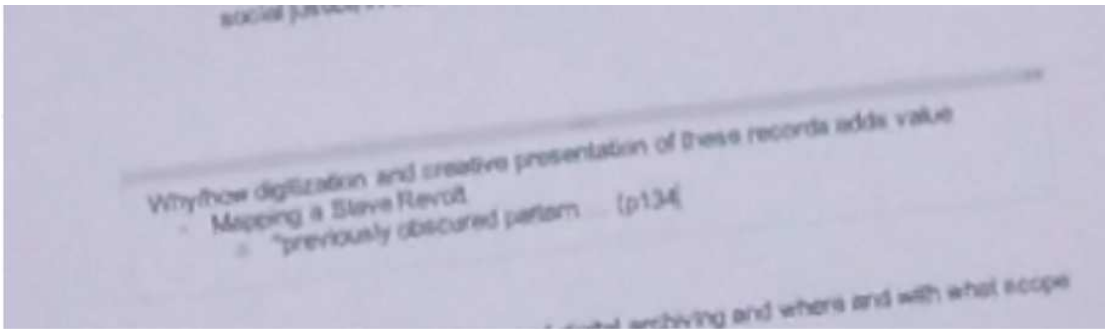


Figure 12 Annotation with page number from P3

In this example, P3 has written a quote of a concept in OneNote and at the same time written down the page number that the quote is from.

For P4 we observed two instances of context being captured with CBBs. In the example shown in fig. 13, the information "see Fig. 2" is contextual to the CBB which is a claim made by the author and is located within the same capture event (highlight).

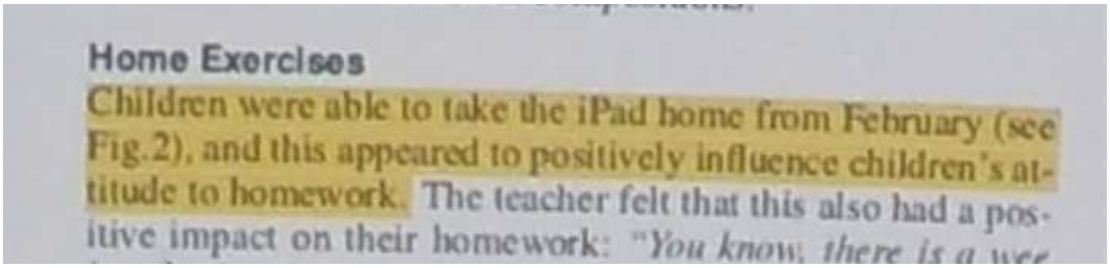


Figure 13 Figure reference number captured in highlight from P4

For P2 we observed 12 instances of integrated context capture. An unnamed feature in LiquidText, which we will refer to as Liquid Capture, automatically captures the following contextual information when text is highlighted and dragged from the pdf view to the canvas section of the interface: file name and page number of the captured information. In P2's session they captured 10 CBBs using the LiquidText capture feature. Below is an example of the object created by a Liquid Capture.

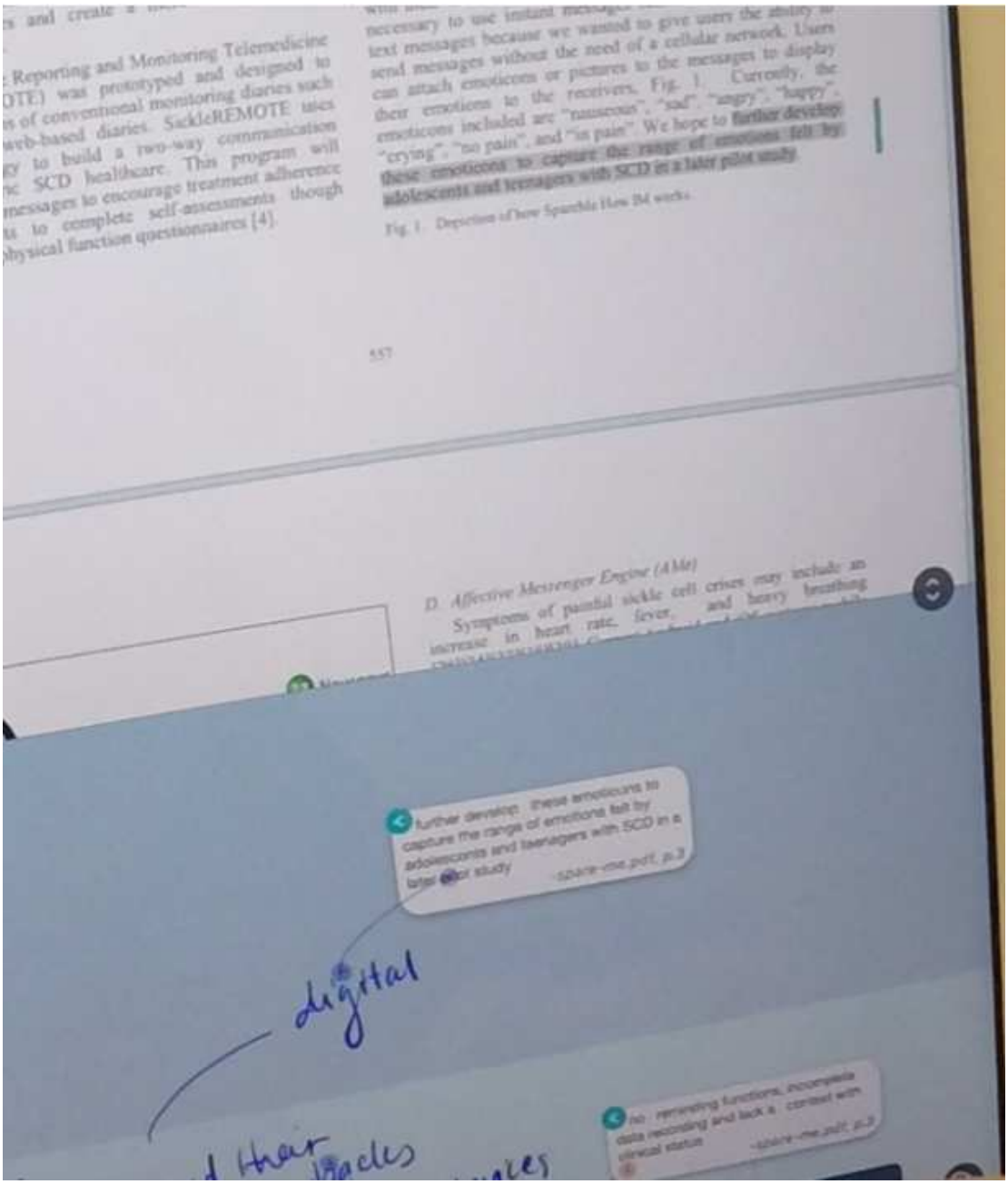


Figure 14 Liquid capture from P2

Located at the center of the canvas in the white text bubble is a CBB captured with the Liquid Capture feature. The file name and page number of the captured context are located at the bottom right of the text bubble.

3.2.2 Standalone Context Capture

Another way that people capture context is separately from any conceptual building block. Examples of this are shown below.

P1 had four capture events which only captured contextual information. In the example below, the highlight located in the middle of the page has been added to the "Preconditions" tag folder using the NVivo tag method.

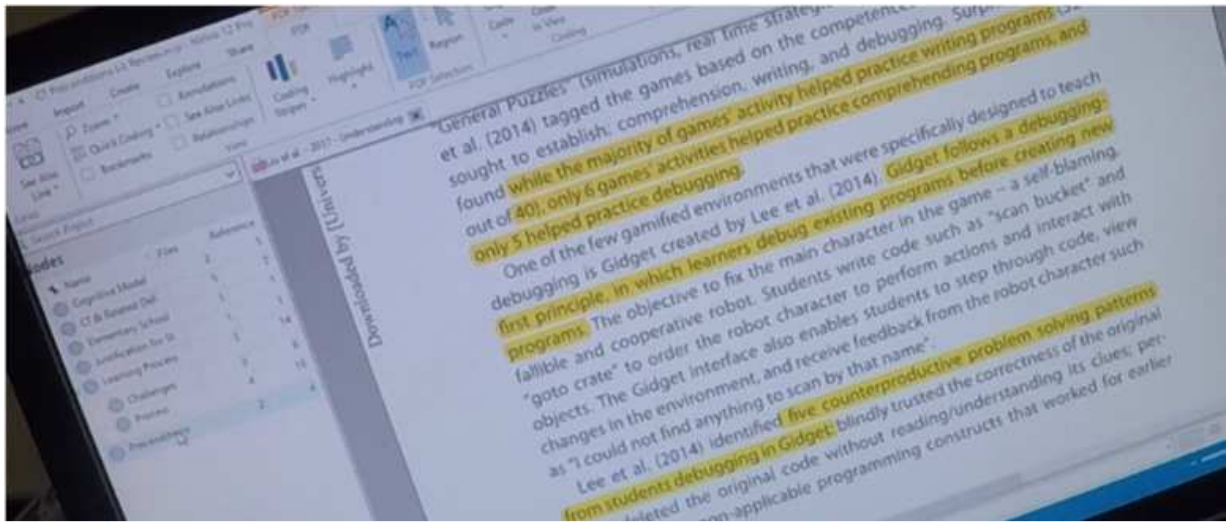


Figure 15 NVivo Tag methodology information

The content of the text is a description of a website, Gidget. Simply as a description, this content is not directly an idea or concept that would be used in a synthesis, but it is information that might support the participants creation or use of another conceptual building block somewhere down the line. For this reason, we have identified it as an explicit context capture.

P3 has two examples of capturing context separately from any CBBs and both are capturing metadata. In figure 11, P3 writes the article title and author of the paper they are reading in anticipation of needing it in the future. There is no CBB attached at the time of the information capture.

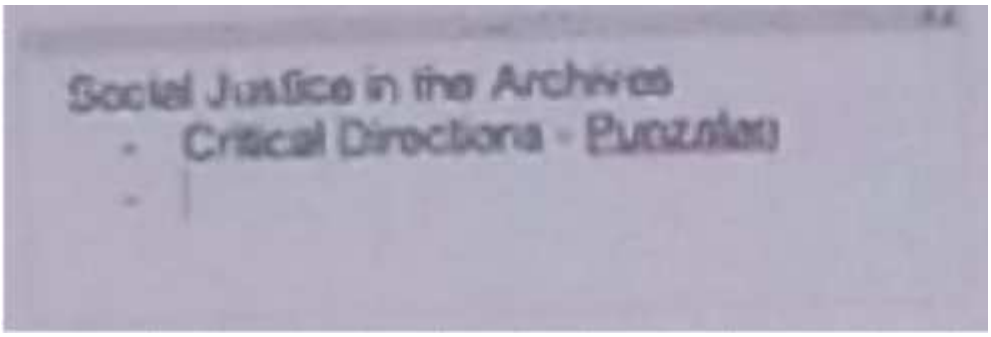


Figure 16 Article title and authors last name from P3

For P4, we observed four examples of explicitly capturing context, all of which were of information related to methodology. Figure 17 shows background information for the study that P4 is reading about.

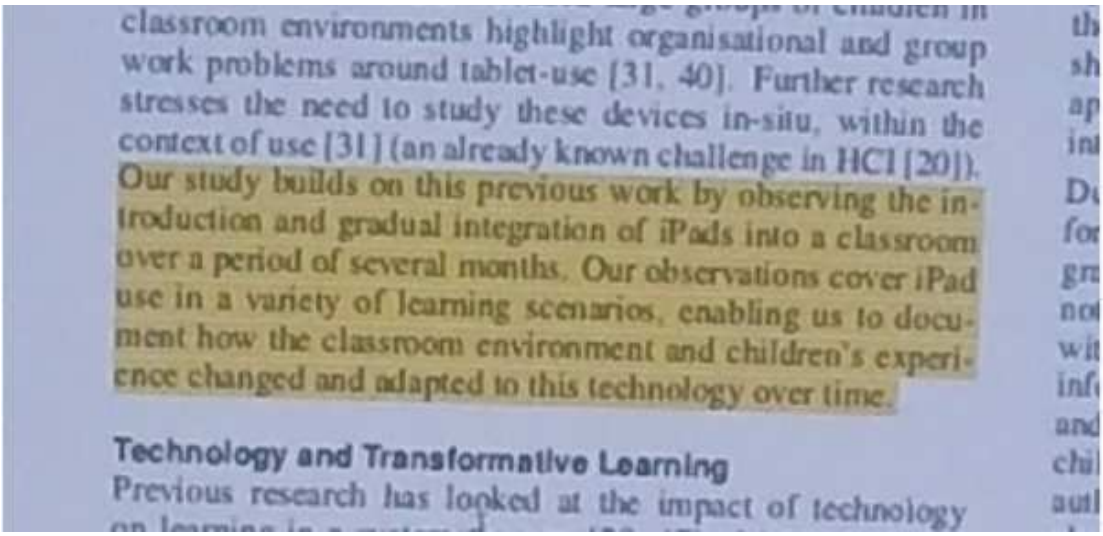


Figure 17 Background information for study from P4

For P2 we observed seven instances of standalone context capture, five of which captured background information/methodological information from the study. In figure 18, the contextual information in the about the design of the study which the participant is researching. The main takeaway here is that all these information capture examples are only of context. No conceptual building blocks are attached.

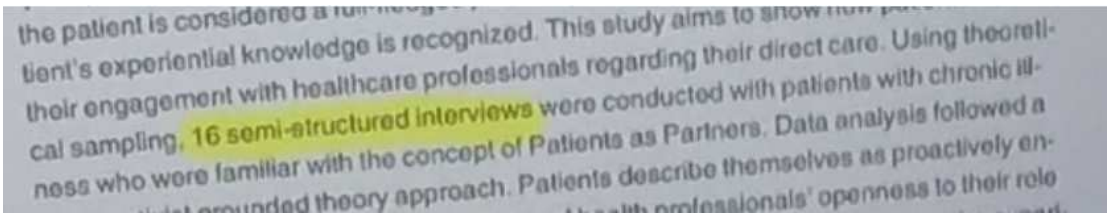


Figure 18 Information about design of study captured by P2

3.2.3 Conceptual Building Blocks as Context

Our last context related observation was that in some instances CBBs served a second role as context for other CBBs.

We observed 17 instances of this for P1, who used the Annotate button in NVivo which automatically creates a link between the highlighted area (claim) and the question that is written in the pop-up text input field. This link allows these two components to act as context for each other. Below is an example of this.

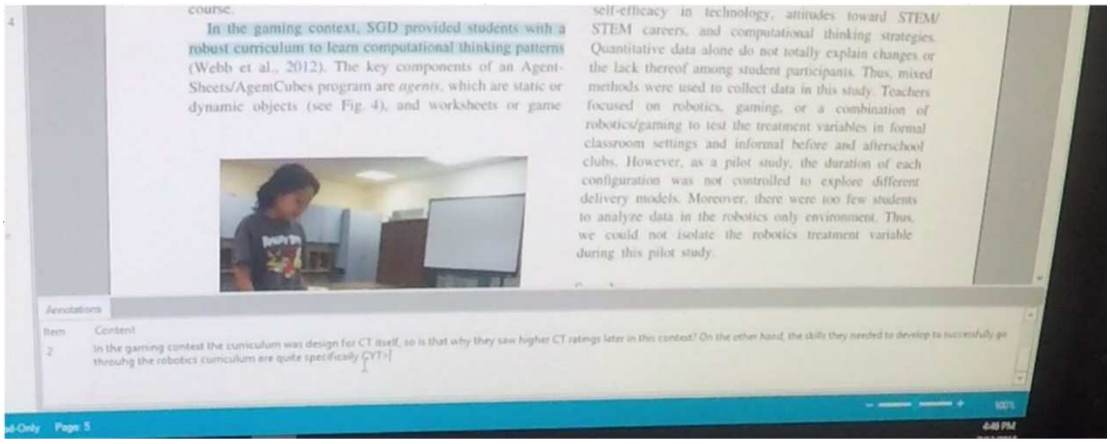


Figure 19 NVivo Annotate button

For P3 we recorded 15 instances of CBBs acting as context for other CBBs. These instances were both margin notes attached to highlights.

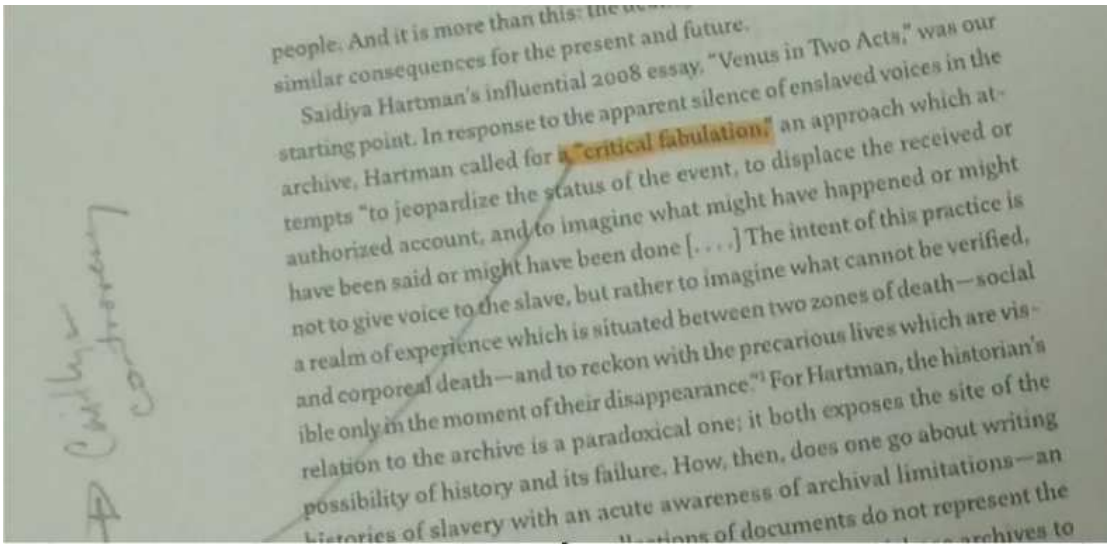


Figure 20 Margin note attached to orange highlight from P3

In the example, we can see that P3 has highlighted "a critical fabulation," a concept, and connected with a line drawn in pencil to the margin note "Caitlin controversy," which is also a concept.

For P4, all of their capture events, except for one, were highlights and no margin notes were taken. The one capture event that was not a highlight was their paper summary and that could be considered an edge case of CBBs serving as context for each other. P4's summary was created by scanning through all of their highlights and including whichever ones they thought were important.

We observed 36 instances of CBBs as context for other CBBs in P2's session. This is due to the structure of LiquidText. LiquidText makes it easy and possible for information to be disembedded from its source, moved around, and manipulated in a canvas view while simultaneously being able to view a captured information from a pdf. This explains why 36 of P2's capture events are of CBB's which are also acting as context for other CBBs.

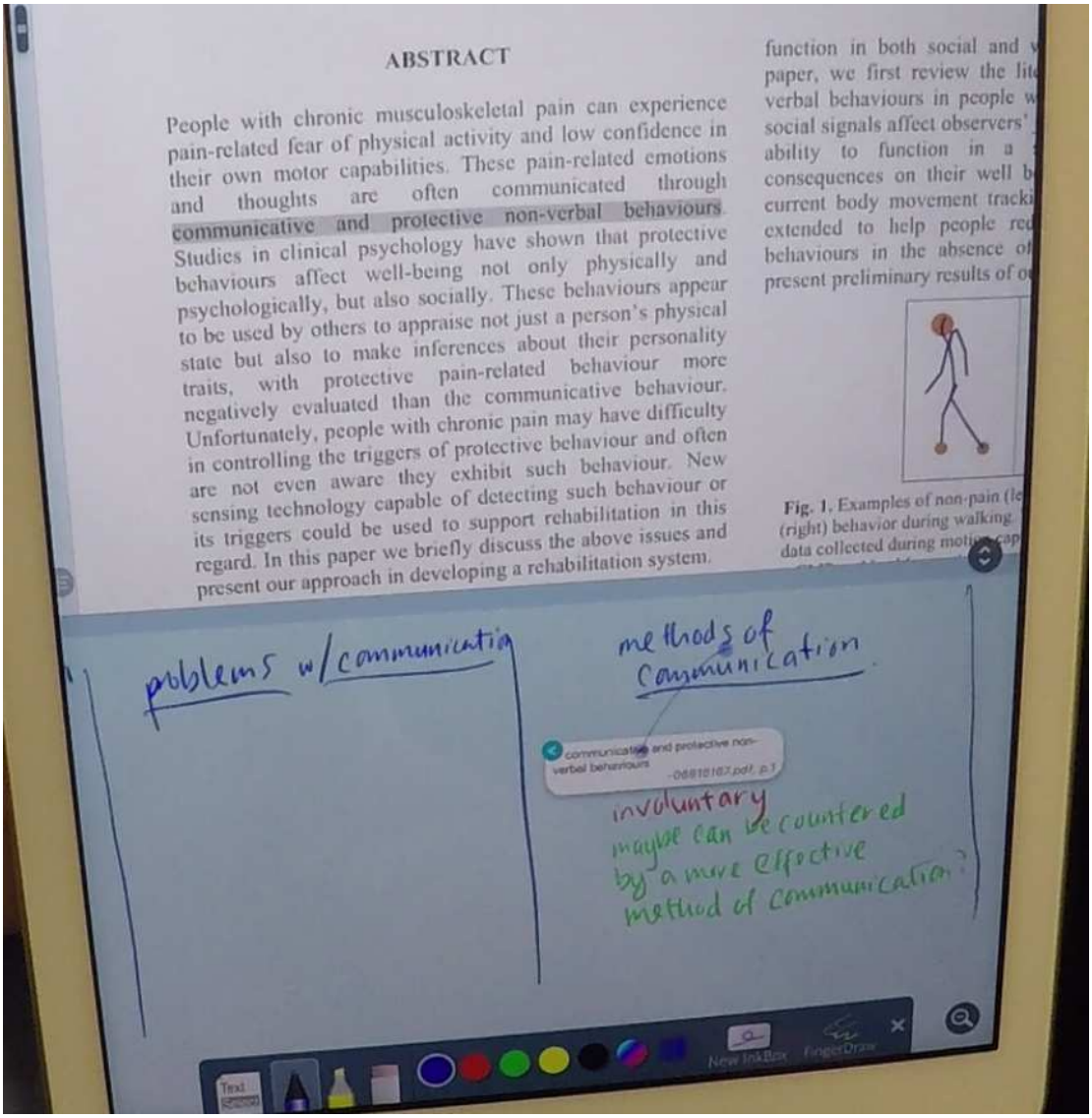


Figure 21 Group of conceptual building blocks on canvas of LiquidText from P2

In figure 21 we can see four separate conceptual building blocks: 1) the blue text header "methods of communication" is a theme being explored, 2) "communicative and protective nonverbal behaviors" in the liquid capture is an idea, 3) "involuntary" in the red text is a comment, 4) "maybe can be countered by a more effective method of communication?" in the green text is a question. Because of their proximity, these conceptual building blocks are now acting as context for each other which influences their interpretation.

3.3 Variations in Context Capture across Tooling Affordances

Next, we compared the two groups of participants with respect to their information capture tooling and behaviors as they relate to the patterns we observed. Below are some of our primary takeaways.

3.3.1 Specialized Group Captures more Context with Conceptual Building Blocks

Our first takeaway is that the specialized group has many more instances of context being captured with CBBs. This is due to the features of each respective participants tooling. For P1, CBBs captured with the tag feature in NVivo can be fully recontextualized in their tag folder using NVivo's Coding Context menu. This functionality is demonstrated in figures 22-25. Figure 22 shows the coding context menu featuring options to show no context, narrow context, broad context, custom, and the entire file that the information is taken from.

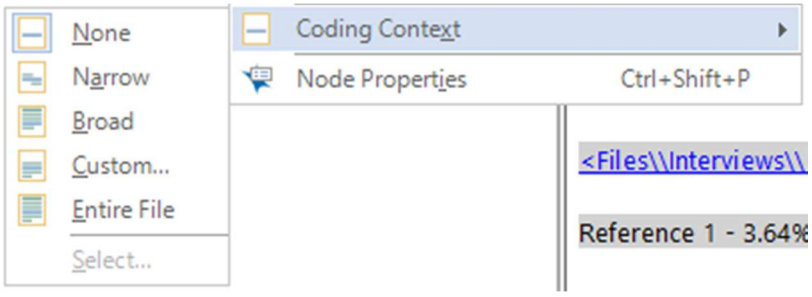


Figure 22 NVivo coding context menu options

Figure 23 shows what information that is captured looks like when no additional context is shown.

<Files\\Interviews\\Charles> - § 1 reference coded [1.84% Coverage]

Reference 1 - 1.84% Coverage

And what has always protected Down East has been the fact that the soil down here is so bad or that the water table is so high that you couldn't ever get it to perk. So you couldn't put a septic tank in.

Figure 23 'None' context option selected

Figure 24 shows what an information capture looks like when the narrow option is selected. You can see some of the text in the proximal area around the information capture is beginning to show.

<Files\\Interviews\\Charles> - § 1 reference coded [1.84% Coverage]

Reference 1 - 1.84% Coverage

threat of the coming development. And what has always protected Down East has been the fact that the soil down here is so bad or that the water table is so high that you couldn't ever get it to perk. So you couldn't put a septic tank in. Well, the state and particularly

Figure 25 'Narrow' context option selected

Then in figure 26, even more of the proximal context area is shown with the broad option selected from the coding context menu.

<Files\\Interviews\\Charles> - § 1 reference coded [1.84% Coverage]

Reference 1 - 1.84% Coverage

Well that's what, that's what got me off the couch so to speak was the threat of the coming development. **And what has always protected Down East has been the fact that the soil down here is so bad or that the water table is so high that you couldn't ever get it to perk. So you couldn't put a septic tank in.** Well, the state and particularly the county government has endorsed the notion of private-package treatment plans. And so if you have a large enough property where you can get the density, that you could make the numbers work to go out and purchase one of these private package treatment plants to treat the sewage – that changed the ballgame cause all of a sudden you didn't need septic tanks. You could put a sewer plant in, treat it and then all you gotta do is dispose of the "clean water". So all of a sudden here's this tract, it's unzoned virtually, the entire part, all of Down East is unzoned. Now you advertise in home builder magazines "waterfront property, no zoning" and see what happens. And this was back six years ago.

Figure 26 'Broad' context option selected

For P2, all information captures in LiquidText show the file name and page number (metadata) of the captured CBB. This is shown below in figure 27.

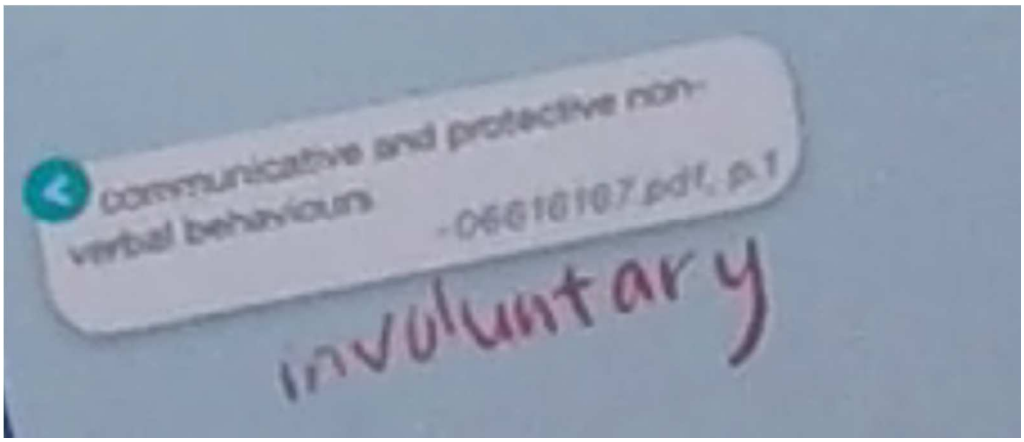


Figure 27 PDF file name and page number of excerpt (metadata) shown in lower right corner of capture bubble

By contrast, we see that there are seven total instances of context being captured simultaneously with a CBB for the generic group (see table 3). Within these examples the contextual information that is captured is page number and paper title, a paper title, and a figure reference number, and several citation numbers. This is all contextual information (metadata) that is either captured automatically or can be easily accessed in the software of the specialized group.

Participant	Number	Example (s)
P1 (S)	11	all are NVivo Tag, one of the 11 also captures in-text citation
P2 (S)	12	Liquid Capture
P3 (G)	5	page number and paper title written with CBB, paper title
P4 (G)	2	figure reference number, metadata of paper for summary

Table 3 Number of conceptual building blocks captured with context (S: specialized group, G: generic group)

It is also the case that the generic group is capturing far more CBBs without any context when compared to the specialized group (see table 4). This means that reusing and interpreting this information later may prove to be challenging for the generic group. As mentioned above,

the specialized group uses tools which can automatically capture metadata and/or allow for the original context to be revisited with very low effort.

Participant	Number	Description
P1 (S)	7	Using the annotate feature to highlight text in NVivo does not capture any context
P2 (S)	0	LiquidText captures context for all events of this participant
P3 (G)	19	No capture methods automatically capture context
P4 (G)	23	Primary capture method (highlight) does not capture context

Table 4 Number of conceptual building blocks captured without context (S: specialized group, G: generic group)

3.3.2 Participants in the Generic group captured qualitatively different kinds of context

For standalone context capture, the specialized group had marginally more instances of capturing contextual information separately from CBBs (see table 5). The main difference here is in the type of contextual information that was captured. P3 in the generic group had two events capturing metadata (author's name and article title, authors name), while the explicit context capture events for other participants were capturing methodological data. The interesting thing to note here is that the specialized group did not capture any metadata manually or explicitly while the generic group did and this is likely due to the fact that their tooling can capture metadata automatically and that they can easily recontextualize. That said, the qualitative differences between the two groups could be due to idiosyncrasies of the researchers perhaps because of their domain or level of expertise.

Participant	Number	Description
P1 (S)	4	methodology, research questions, methods
P2 (S)	7	characteristic of population in study, background info for study, background info for study, background info for study, background info for study
P3 (G)	4	article title and author last name, author, background info

P4 (G)	4	background info, methodology/background info, methodology (2)

Table 5 Standalone context capture (S: specialized group, G: generic group)

3.3.3 Specialized Group has more examples of Conceptual Building Blocks acting as Context

Far more instances of CBBs acting as context occur in the specialized group (see table 6) because their tooling allows for the CBBs to be disembedded and viewed in new locations more seamlessly than the tooling of the generic group. Embedding these instances in new locations can facilitate new interpretations of that information. We observed 17 instances of CBBs as context for other CBBs for P1 because of their ability to easily tag and send annotations to folders in NVivo where they can be categorized and viewed among other captured information. In P2's session we observed 36 instances of this because of the information mobility afforded by LiquidText's canvas where information can be sent and arranged.

Participant	Number	Description
P1 (S)	17	from the annotate feature (This is an automatic version of P2's CBBs as context events) which allows a note to be added to a highlighted area + NVivo tag folder which holds many CBBs in a new location
P2 (S)	36	canvas area allows for easy maneuvering of CBBs into different locations aka to be context for other CBBs
P3 (G)	15	this is when they write a margin note that they attach to a highlight. Manual version of the NVivo Annotate button from P1
P4 (G)	1	No examples of CBBs as context for each other except maybe the summary made at the end. The summary (miniature synthesis) could be conceptually similar analogue to liquid text canvas area and NVivo tag folder

Table 6 Conceptual building blocks as context

On the other hand, the automatic disembedding and embedding process for information is either not as seamless or not available for the generic group. The notes taken by P3 on physical paper need to be manually typed up in a new location, in their case, OneNote. Once in OneNote

they can use copy and paste to move their information around in OneNote's canvas view. All instances of CBBs acting as context were either margin notes attached to a highlight on printed paper or notes arranged near each other in OneNote. The more deliberate, constrained nature of this information capture workflow may account for why P3 only had 15 instances of this.

There is an interesting observation here about the usage of OneNote's canvas view when compared to P2's usage of LiquidText's canvas view. P3 uses OneNote's canvas more like a notebook or word document, writing in outline form whereas P2 organizes pieces of information on the LiquidText canvas and uses the positioning of the notes to associate notes together.
For P4 we observed very little information mobility in their workflow. The information they highlighted in the Apple Pdf reader was occasionally copy and pasted over to google docs, but for the most part the summary that they wrote was done manually while looking at their highlights.

Chapter 4: Discussion and Conclusions

4.1 Summary and discussion of findings

In this study we sought to understand researchers' interactions with contextual information in the literature review process. In response to our first research question: how are people describing context with the information that they capture and create for synthesis, we observed three groupings of context related interactions by our participants:

1. Capturing context while capturing conceptual building blocks
2. Capturing contextual information separately from conceptual building blocks
3. Conceptual building blocks becoming context for other conceptual building blocks

By conducting a protocol analysis of the four researchers divided into two groups (group 1 using specialized tools and group 2 using generic tools) we discovered distinct differences between the groups in their patterns of context capture. In our second research question we asked: how do patterns of context capture vary across literature review tools. What we found was that there was more and richer context capture with the group using specialized tooling and this was due to the affordances of their tools.

First, the group using more specialized tools captured more contextual information per conceptual building block. With P1's tool Nvivo, the tagging feature captures information, adds it to a tag folder, and attaches the file name (contextual metadata) of where the information is being captured from. P2's tool, LiquidText, takes this a step further by attaching the file name as well as the page number to information captured through its liquid capture feature.

These affordances led to our second observation which is that the specialized group captured qualitatively different kinds of context, specifically they did not capture any metadata on their own. By contrast, we observed P3 from the generic group capture the following metadata manually: article titles and author names.

The last observation is that the tools of the specialized group afford a degree of mobility to the information that they capture which allows conceptual building blocks to be positioned in novel ways. In NVivo this means being able to easily organize information objects into folders and in LiquidText this means being able to position them on a digital canvas.

4.2 Limitations

The following limitations should be considered when interpreting or generalizing from these findings:

The data collection part of the study involved three sessions (guided tour, information capture, and information reuse) with ten participants but we chose to analyze only four of the participants and only the information capture sessions. Both concessions were made to scope down the analysis to a level where we could go in depth into the participants' interactions with contextual information. We were able to reach some degree of saturation from analyzing the 4 participants split into the two groups, but analyzing the full 10 might have strengthened the observed patterns that emerged.

As the sample size is relatively small, some caution should be exercised when generalizing to others. Some of the participants' behaviors may be idiosyncratic in addition to the fact that there is and can be considerable variability in workflows through the literature reviewing process. There is variability in which tools are selected and how they are used, and this can vary by domain or profession as well.

The omission of analysis of the information reuse sessions was also due to scope and time constraints. The reuse sessions show the information behaviors of the researchers reusing the information that they captured in the earlier session and will complete the picture of the synthesis cycle in the literature review process. They would create a clearer picture of what context looks like because, in theory, we would see which contextual information was needed for information reuse.

Another limitation due to scope and time constraints was the omission of analysis of ephemeral information capture events. The addition of ephemeral information capture events would give a more holistic view of the information capture process by providing some insight into the information that is recognized by participants, but not committed to durable actions. A potential observation along this train of thought might be if a researcher takes a mental note and

is unable to recall it in the information reuse session. A line of inquiry could then identify possible motivations, pain points, and design solutions related to that interaction.

4.3 Design Implications

Returning to the main purpose of this study of informing the design of literature reviewing tools, we now discuss three main design implications.

First, tools that want to support context should automatically capture metadata during information capture. Automatically grabbing the authors name, page number of a quote, or the title of a paper that information is taken from makes information reuse much easier. Participants in the group using tools that support literature reviewing did not spend any time capturing metadata because their tools captured it for them. The most seamless example of this was LiquidText, pictured in figure 28, where every piece of excerpted text had the file name and page number automatically attached to it. If metadata is captured and easily available this can save a considerable amount of time for researchers.

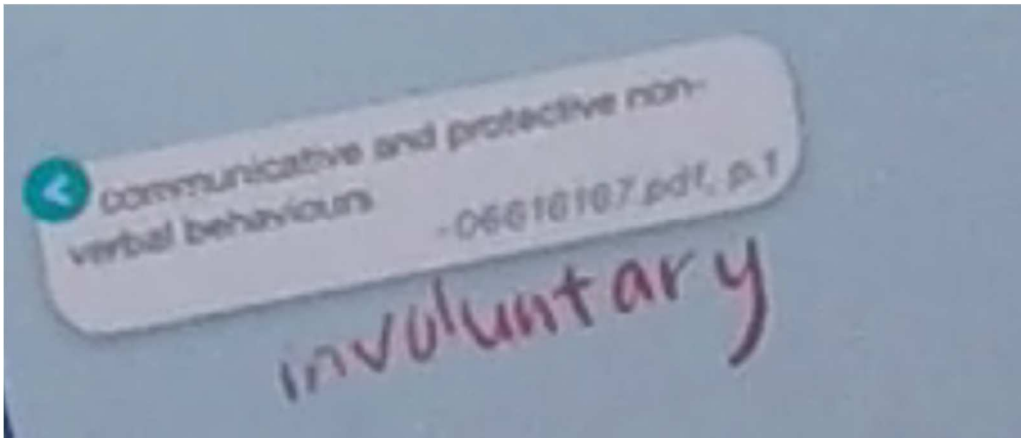


Figure 28 Automatic metadata capture in LiquidText

Next, tools should support more effortless recontextualization to make it easy to view and/or return to the original context of information. One way to do this is a method called transclusion where information is disembedded from one place, embedded in another, and able to

be referenced back to its source from the new location. Information captures in LiquidText are examples of this. We saw another useful method of recontextualization in NVivo's coding context menu. For a captured piece of information, the user is able to reveal varying degrees of the information that surrounds it at its source location.

We came across evidence that more effortless recontextualization might be useful after repeated observations of metadata located around conceptual building blocks in their source location. An example of this is figure 29 where you can see a page number located right outside of the area defined by the highlight.

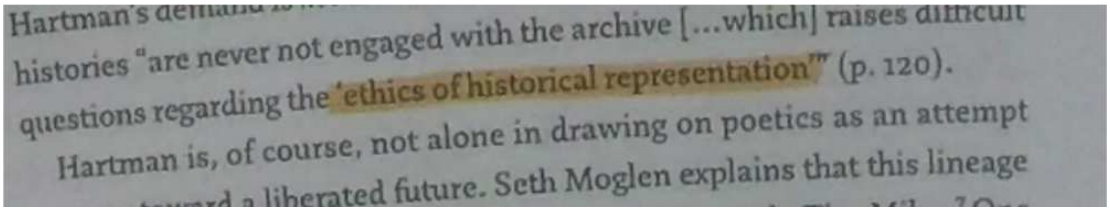


Figure 29 Page number situated to the right of a conceptual building block outside of highlighted area

We witnessed seven instances of metadata located outside of a captured conceptual building block for both the group using specialized tools and the group using generic tools. What is important to note is that the group using the specialized tooling has easier access to this information. P1 using NVivo can progressively reveal the area surrounding a conceptual building block that they have captured through the coding context menu. The information captured in LiquidText is linked back to its source so that one click will show the user the original context of the conceptual building block. The tools used by the unsupported group do not feature affordances like these which makes recontextualization a manual, more labor-intensive task.

Finally, tools should support information mobility. Information mobility is the ability and degree of ease for information to be moved around. Less mobility typically means more effort to move information from one place to the next which can result in loss of contextual information.

Physical notes on paper require more effort to move around when compared to copy and paste in the digital space. In LiquidText we see the most flexible form of information mobility because you can move information around on a digital canvas. We also saw a form of this in NVivo where information could be put into tag folders. Giving information more mobility can lead to more ways for conceptual building blocks to become context for each other. This is important for synthesis because it allows information to be compared/contrasted/modified/accreted more easily.

4.4 Future Work

The next logical progression from this study, which looked entirely at information capture, is to look at information reuse sessions so that a fuller more accurate depiction of the information flows can be observed. Observing the information reuse sessions would also clarify which of the captured information truly was a conceptual building block because we would see which of the information was ultimately used. Furthermore, in observing which information is reused we would see which contextual information that was captured was used as well. Or, we might see which contextual information was needed but not captured. There is no guarantee that the researchers will reuse all the information from these capture sessions, so we would want to follow them over multiple sessions to have a better chance of observing what is happening.

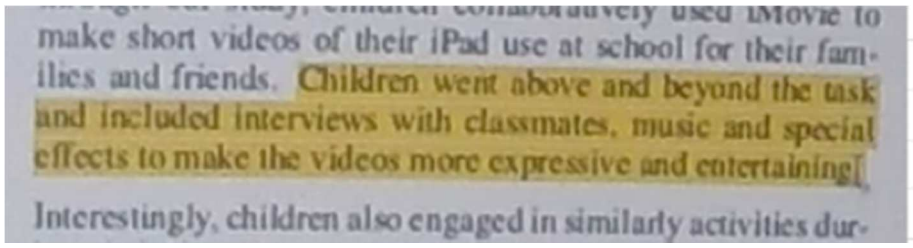
Another topic worth exploring is information trajectory, or more specifically, how much information should be captured? We saw earlier in Anderson et al. (2008) how trying to predict which information a caretaker would need in a drop-down menu was not possible. Instead of the information capturer deciding which information might be necessary in the future, it should be the information reuser making that decision. By this logic it may make the most sense to capture

as much information as possible in the information capture stages, then have appropriate, strategized information retrieval methods available for the reusers so they are not overwhelmed. In our study we observed that the specialized tooling group captured more information because of their tooling affordances. This may make save them time in the information capture stage and potentially make information retrieval and reuse easier, but with these may come tradeoffs that we have not explored.

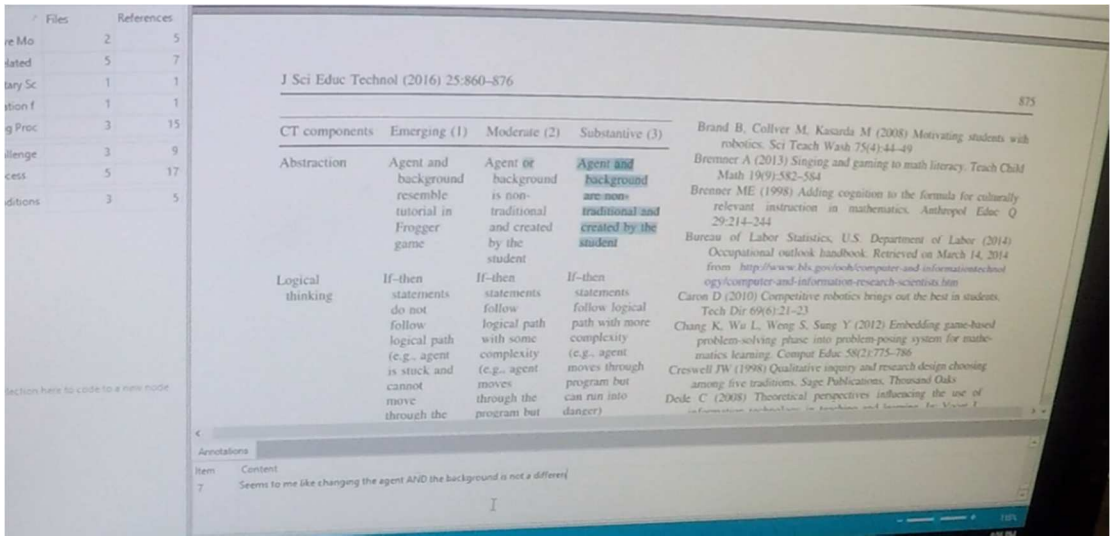
Another interesting topic to consider next is the idea of "desirable difficulty" or "frictionful" design. Throughout the paper we observed how many of the affordances of the specialized tools can make the literature reviewing process more seamless. On one hand, capturing too much contextual information may lead to information overload in the information reuse stage. On the other hand, not capturing enough or no contextual information may make interpretation in the reuse sessions challenging. This concept is also suggested as an interesting next step for future work in Qian, Fenlon, Lutters, and Chan (2020) where they describe it as "desirable difficulty." What is the right balance between having a frictionful and frictionless workflow? If the literature reviewer is under time pressure and has many documents to process, then the benefits of having friction will probably be outweighed by the cost of them. One potential argument for more frictionful design is that working more slowly and deliberately may help create more deep, lasting knowledge. If you use generic tooling, then you will not have the option to decide what your desirable difficulty is. Specialized tools give you that option, so having power features available when you need them might be the way to go. Our suggestion is to give the user a choice because it might vary depending on the situation.

Appendices

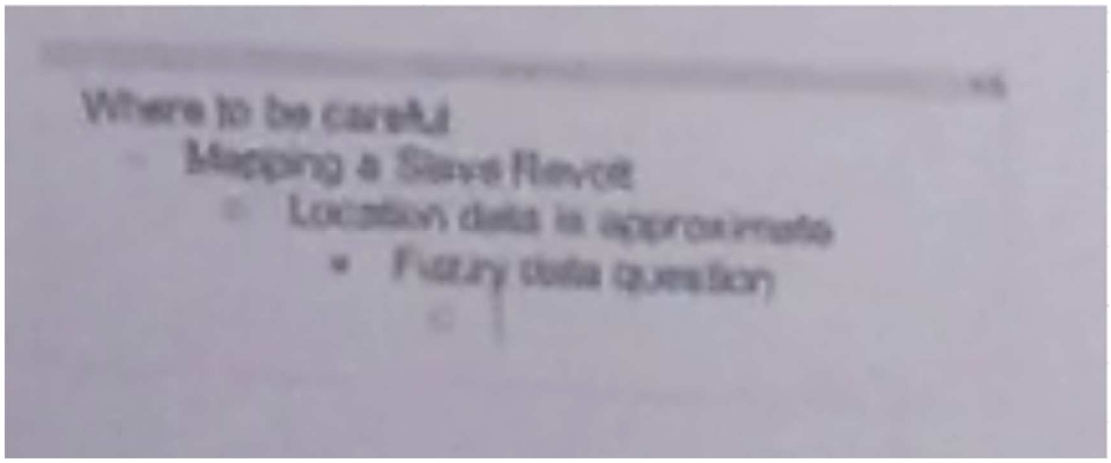
Verbal Cues for Conceptual Building Blocks



"I feel like we may have seen this a little bit" – P3 relating an observation from their literature review to their own work



"I have an issue here" – P1 disputing a finding



“I’ve got another piece to this that have somewhere” – P2 suggesting that this is a concept that they can add to

References

Ackerman, M. S., & Halverson, C. (2004). Organizational memory as objects, processes, and trajectories: An examination of organizational memory in use. *Computer Supported Cooperative Work (CSCW)*, *13*(2), 155-189.

Anderson, S., Hardstone, G., Procter, R., & Williams, R. (2008). Down in the (data) base (ment): Supporting configuration in organizational information systems. In *Resources, co-evolution and artifacts* (pp. 221-253). Springer, London.

Dourish, P., Bellotti, V., Mackay, W., & Ma, C. Y. (1993, December). Information and context: lessons from the study of two shared information systems. In *Proceedings of the conference on Organizational computing systems* (pp. 42-51).

Dourish, P. (2004). What we talk about when we talk about context. *Personal and ubiquitous computing*, *8*(1), 19-30.

Ervin, A. M. (2008). Motivating authors to update systematic reviews: practical strategies from a behavioural science perspective. *Paediatric and perinatal epidemiology*, *22*, 33-37.

Hinrichs, J., Pipek, V., & Wulf, V. (2005). Context grabbing: assigning metadata in large document collections. In *ECSCW 2005* (pp. 367-386). Springer, Dordrecht.

Holbrook, A. (2007). 'Levels' of success in the use of the literature in a doctorate. *South African Journal of Higher Education*, *21*(8), 1020-1041.

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?. *The Review of Economic Studies*, *76*(1), 283-317.

Knight, I. A., Wilson, M. L., Brailsford, D. F., & Milic-Frayling, N. (2019, March). Enslaved to the Trapped Data: A Cognitive Work Analysis of Medical Systematic Reviews. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval* (pp. 203-212).

Lutters, W. G., & Ackerman, M. S. (2007). Beyond boundary objects: collaborative reuse in aircraft technical support. *Computer Supported Cooperative Work (CSCW)*, *16*(3), 341-372.

O'Hara, K., Smith, F., Newman, W., & Sellen, A. (1998, January). Student readers' use of library documents: implications for library technologies. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 233-240).

Petrosino, A. (1999). Lead authors of cochrane reviews: Survey results. *Report to the Campbell Collaboration. Cambridge, MA: University of Pennsylvania.*

Qian, X., Fenlon, K. Lutters, W. G., & Chan, J. (2020) Opening Up the Black Box of Scholarly Synthesis: Intermediate Products, Processes, and Tools. *Proceedings of ASIST 2020*. Proceedings of ASIST 2020.

Qian, X., Erhart, M. J., Kittur, A., Lutters, W. G., & Chan, J. (2019, November). Beyond iTunes for Papers: Redefining the Unit of Interaction in Literature Review Tools. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (pp. 341-346).