Comprehending Spreadsheets: Which Strategies do Users Apply?

Karin Hodnigg, Christian Macho, Martin Pinzger
Software Engineering Research Group
University of Klagenfurt
Klagenfurt, Austria
firstname.lastname@aau.at

Dietmar Jannach
Information Systems Research Group
University of Klagenfurt
Klagenfurt, Austria
dietmar.jannach@aau.at

***Abstract*—Spreadsheets are a ubiquitous means of communication and decision making, and adequate comprehension is required for reading, interpretation as well as maintenance. Despite being around for forty years, only selected aspects of comprehension were studied, mainly focusing on error detection and correction, and largely disregarding a more holistic view on spreadsheet comprehension processes.**

In this paper, we provide the first steps towards a deeper understanding of how users approach comprehension through a user study. In this study, we tasked eight spreadsheet users to describe their thought processes while trying to familiarize themselves with a real-world spreadsheet. The transcripts of these eight think-aloud studies were then used to identify 16 comprehension indicators validated by three experts. With these categories, we set out to qualitatively identify frequent comprehension patterns that the participants applied to understand spreadsheets.

Although the comprehension process is highly individual, we could identify frequently occurring patterns, such as assumptions can imply consolidations and realizations, and realizations often happen before contradictions. The results of this work can be used to design better tools that assist users in comprehending spreadsheets.

***Index Terms*—spreadsheet, comprehension, empirical software engineering, controlled experiment, maintenance**

I. Introduction

Spreadsheets as a means of computation [1] and communication [2] are handed around to various experts [3]. Adequate comprehension of a spreadsheet is required during maintenance, audits or transfer of spreadsheets. In particular, the correct interpretation of results relies on the ability to understand the spreadsheet. The logic and documentation, though, is often hidden behind formulas [4] and not made explicit. Comprehending a given spreadsheet program and reading it is thus a main competence of spreadsheet users.

The comprehension process is complex, because the conceptual model that the spreadsheet programmer has in mind, is hard to grasp or visualize due to the lack of abstraction and modularity mechanisms [5]. Studies on short term memory and spreadsheet users have shown that the user's mental model is very dependent on his/her domain [5], [6].

In this paper, we study user comprehension to understand how end users think in order to provide adequate tool support. Current research primarily focuses on selected aspects of comprehension, while other studies did not consider the new features of spreadsheet IDEs. Thus, this paper focuses on the following research question:

Which comprehension strategies do spreadsheet users apply?

We conducted an experiment with 8 participants asking them to make themselves familiar with an unknown spreadsheet. We then manually labeled the respective transcripts with the categories that we defined in our taxonomy of comprehension indicators. To gain a deeper understanding of the comprehension strategies that users apply, we then analyzed the labeled sentences aiming at identifying frequent comprehension patterns. We found that, for example, insights often happen after assumptions and before (new) confusion, or that summarizing usually happens before both, insights and misunderstanding. With the results of this work, we provide the basis for further studies to understand spreadsheet comprehension which can ultimately help to better support spreadsheet users with better and more individualized tooling.

The remainder of the paper is organized as follows: In Section II we situate the paper according to related work. Section III details our idea and presents the preliminary study. Finally, Section IV presents the threats to validity, prospects future work, and concludes the paper.

II. Related Work

Although some spreadsheet comprehension studies and experiments are available, they often focused on selected aspects of comprehension or were conducted with spreadsheet environments that do not provide features that help in understanding spreadsheets, such as dependency highlighting, reference arrows or formula views:

Vemuri *et al.* investigated as early as in 1992, whether data dependency graphs deduced from spreadsheet models aided in spreadsheet maintenance tasks [7]. Hendry and Green conducted a study in 1994 with 10 users to analyze strengths and weaknesses in the user interface [8]. In 2003, Burnett *et al.* studied the use and benefits of assertions in spreadsheets with 59 participants in a controlled experiment [9]. In 2008, Bishop and McDaid conducted an experiment with 13 professionals, on how they detected errors in a spreadsheet by activity tracking [10], [11]. The usefulness of data flow visualization and its impact on knowledge transfer was evaluated by Hermans, Pinzger and van Deursen [12] in 2011. In 2012, 40 professionals assessed the understandability of formulas in a large scale study conducted by Hermans *et al.* [3]. Kohlhase *et al.* studied the impact of context in spreadsheet comprehension with 3 participants in 6 interviews in 2015, concluding that readers miss the context that the authors of the spreadsheet have [2].

These studies and papers were focused on specific tasks, i.e. finding an error, adapting to new requirements, but are lacking a more holistic view of the process of spreadsheet comprehension: How do users approach an unknown spreadsheet, if there is no task to perform other than "mere understanding"?[1]

III. Identifying Comprehension Strategies

We set out to qualitatively study which comprehension strategies spreadsheet users apply when trying to understand unknown spreadsheets aiming to answer our research question *Which comprehension strategies do spreadsheet users apply?*. We conducted an experiment with 26 participants in which we handed them a previously unknown spreadsheet and asked them to think aloud while making sense of the sheet. In this preliminary work, we studied the transcripts of the first 8 participants and manually identified comprehension strategies. We leave the analysis of the remaining 18 transcripts obtained with two further spreadsheets for future work.

A. Study Setup

Participant Selection. The only requirement that participants need to fulfill was to professionally work with spreadsheets (*i.e.*, regularly uses spreadsheets at their work). To that extent, we invited 26 persons to participate in the study of which 8 were included in our preliminary experiment. The eight participants consisted of five female and three male participants from age 23 to 51, stemming from different sectors of employment (Education 37.5%, Information and Consulting 25%, Research 12.5%, Industry 12.5%, Energy 12.5%). Their experience with spreadsheets ranged from 8 to 34 years averaging at 18 years. When asked to assess their expertise, 2 rated themselves as intermediates, 5 as advanced and one as expert.

Spreadsheet Selection. We selected the spreadsheet for this study from the ENRON [14] corpus consisting of more than 15,000 spreadsheets. For all processible spreadsheets in the corpus, we selected spreadsheets that are of reasonable size (100-2,000 formula cells; 20-100 distinct formulas; 500-10,000 cells). Furthermore, we excluded spreadsheets with errors, references to external sources, and procedural extensions (*e.g.*, VBA macros) to avoid unnecessary complexity.

The main author investigated the 114 spreadsheets and selected the 6 sheets that she found most suitable (*i.e.*, for which she anticipated that the domain is unknown to the participants and the participants have never seen this sheet before) and which vary in size. After a pilot experiment with each of the sheets, we approved three that we have used in the experiment. Out of these three, we analyzed the 8 transcripts obtained with the first spreadsheet of which we report the results in this paper.

This spreadsheet contains a calculation to estimate whether - and if, at what rates - an investment into a baseball stadium pays off. It consists of 951 non-empty cells with the following distribution: labels (21%), input cells (18%), computation cells (40%) and output cells (17%). In total, it contains 546 formula cells with 64 distinct formulas.

B. Study Execution

After providing informed consent, each participant was confronted with the unknown spreadsheet and was asked for an initial description of it. The concrete task was *"Please describe briefly the purpose and content of the sheet"*. We did not enforce any time restrictions but let the participant decide when she was content with her insights to answer the question.

Commenting (thinking-aloud) was encouraged, but not enforced during the experiment. However, the interviewer sometimes interposed reflexive questions in the form of "What are you focusing on?" or "Why did you do that?" to evoke remarks. At the end of the task a reflexive question was posed, "How did you achieve this?". By their nature, the in-process statements were more insightful.

We recorded both, voice and screen of the interview, and the resulting transcripts built the basis for the analysis of the spreadsheet comprehension strategies in this paper. At the end of the interview, participants were asked to rate the perceived complexity of the spreadsheet in terms of size, formula complexity and domain complexity.

C. Taxonomy

For analyzing the transcripts, a team of three raters applied a card-sorting approach on a subset of sentences to identify the comprehension indicators. We randomly selected sentences from the transcripts to classify until we reached saturation (*i.e.*, we could not find any new indicator). In our case, we could reach saturation after three iterations (150,100, and 150 sentences, respectively).[2] A sentence can generally contain no or all indicators.

The following 16 comprehension indicators were identified:

- The **Interviewer (IA)** marks sentences that indicate that the interviewer interacted in a way that influenced the participant (e.g. by answering questions).
- **Labels (LL)** marks sentences that reference labels, *i.e.*, textual description in cells.
- **Formula (FF)** marks sentences that refer to functions or formulas.
- **Dependencies (DEP)** marks sentences the reference single cells, ranges, or cell names.
- **Values (VAL)** marks sentences that refer to concrete cell values in the spreadsheet.

[1]Recently, Srinivasa *et al.* [13] also studied spreadsheet comprehension strategies and came to similar results compared to this work.

[2]From all 26 transcripts over all three spreadsheets: Those 400 sentences represent 15% of sentences transcribed.

- The **format (FORM)** marks sentences that mention font styles, colors or borders.
- **Help (HLP)** marks sentences that refer to the use of either internal or external help (e.g. search engines).
- **Structure (STR)** marks sentences that reference tables or blocks of cells.
- **Interpretations (ITP)** marks sentences that denote data-driven readings of the spreadsheet construed by the participant; to the effect of ”that means”.
- **Assumption (AS)** marks sentences that describe suppositions without full or only partial information; to the effect of ”I believe”.
- **Question (QUES)** marks sentences that describe rhetorical and actual inquiries by the participant.
- **Consolidation (CONS)** marks sentences that refer to aggregating, summarizing or recapping information.
- **Realization (REAL)** marks sentences that describe the concrete insights.
- **Contradiction (CONTR)** marks sentences that denote contradictions of prior statements.
- **Incomprehension (INC)** marks sentences that denote a lack of understanding.
- **Micro-Strategies (TACT)** marks sentences that subsume all statements that describe the participant’s course of action.

Armed with the 16 indicators, the three raters together manually labeled the sentences of the transcripts of the first 8 participants. Sentences with conflicting labels were discussed until consensus was reached. Table I shows several examples of labeled sentences.

TABLE I: Examples of labeled sentences.

Exp.	**Indicator**	**Sentence**
009	REAL, CONS	Oh, that’s cool - ah! - this is DiscRate, the discount rates row containing 5% #REAL #CONS
010	TACT	So I’ll just start at the very top left #TACT
010	LL, AS, INC	It says something about home matches #LL, I mean - maybe it’s about a stadium #AS, I have no idea #INC; It’s actually very confusing.
019	VAL, AS	It says $330.5 #VAL, that would have to be kind of thousands or millions #AS.
019	REAL, DEP	These - the numbers - well, 339 #VAL, okay, that’s clear that that’s referring to up there #REAL #DEP.
020	FF, AS	So you’re discounting that apparently, over the years #FF #AS.
023	HLP	Hmm, okay, [Opens search engine.] #HLP).
023	AS, ITP, QUES	I mean, the only thing I could think of is the population - how many are living there #AS; It looks like that that’s steadily increasing #ITP - but why is there 2000 in between #QUES?
028	CONTR	Let’s see if that’s right [enters formula ’=D6-13%’ in O16] Oh well, no #CONTR.

Figure 1 shows the frequency of the comprehension indicators. We see that the most frequent indicator is Consolidation followed by Labels and Incomprehension. The indicator Help, for example, is less frequent. We also found that participants used external help, in particular Google Search, more frequently than the internal help provided by the spreadsheet IDE.

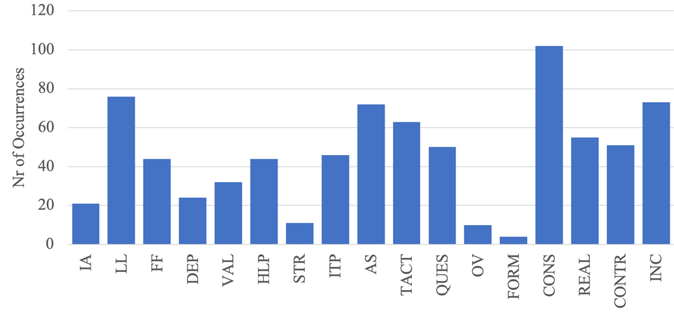

Fig. 1: Comprehension indicator frequencies.

D. Identifying Comprehension Patterns

Using the annotated transcripts, we leveraged the natural order of the labels of the sentences in the transcripts to identify comprehension patterns. Figure 2 shows an example of such a pattern.

{CONTR}, {HLP}, {*REAL* CONS FF}

Fig. 2: Example pattern.

We see that the participant found conflicting information (CONTR), chose to use help (HLP), which helped her to formulate the insight (REAL, CONS, FF).

As comprehension patterns are represented by sequences of our comprehension indicators, we aimed at leveraging the well-known cSpade algorithm [15] to extract sequential patterns. We experimented with several configurations and combinations of input but the algorithm could not find outstanding patterns. It yielded more than 80,000 patterns that only were from minor statistical relevance – the lift of the rules was 8 at most for only a few patterns. Hence, we discontinued this analysis and decided to manually analyze the labeled data.

For each occurrence of an indicator in a sequence, we counted the number of occurrences of any indicator in an range of 5 steps backwards and 3 steps forward. For example, for a the sequence in Figure 2, we count 1 for HLP and 1 for CONTR, respectively, given REAL as a starting point. With this, we counted the number of occurrences of each comprehension indicator in the context of another indicator. We used 5 and 3 as thresholds because we aim at looking further in the past to analyze which indicators happen after which other indicators.

For every comprehension indicator in the sequences, we also computed the relative frequency of each occurring indicator. For example, if we found 175 total labels in the backwards step and 34 of them are from the type Labels (LL), the relative frequency is calculated with $34/175 = 0.19$. Table II shows the results of the relative frequencies for selected indicators. We can see that, for example, insights often happen after assumptions and are often followed by (new) confusion (first row).

The concepts of understanding and confusion are of primary concern during the comprehension process. For *understanding*, we looked into the indicators realization (REAL) and consolidation (CONS), for *confusion*, we considered incomprehension (INC) and contradiction (CONTR). The results in Table II

TABLE II: Predecessors and successors for selected indicators.

Indicator	Predecessor	Successor	Interpretation
REAL	AS (0.14) LL (0.11) CONS (0.10)	CONTR (0.16) INC (0.14) FF (0.09)	Realizations follow assumptions and happen before (new) confusion.
CONS	LL (0.15) AS (0.12) ITP (0.10)	INC (0.18) REAL (0.13) AS (0.12)	Consolidations either happen before insights or misunderstanding.
CONTR	AS (0.11) REAL (0.11)	AS (0.14) INC (0.13) FF (0.10)	Assumptions can enclose contradiction, and incomprehension can follow it.
INC	LL (0.11) AS (0.11) TACT (0.09)	ITP (0.12)	Incomprehension often follows assumptions and labels, and users tend to interpret afterwards.
HLP	TACT (0.13) INC (0.12) REAL (0.11)	TACT (0.17) INC (0.17) CONTR (0.15)	Using help is often explicitly expressed, often following either insights or incomprehension. Not always does it help.
LL	AS (0.14) ITP (0.11) INC (0.11)	AS (0.17) QUES (0.12) ITP (0.11)	Labels provoke users to speculate, but are also used in interpretation.
FF	LL (0.19) REAL (0.12) AS (0.09)	AS (0.16) TACT (0.11) QUES (0.11)	Formulas can be preceded by realizations, and followed by assumptions or thinking about next steps.
VAL	LL (0.23) AS (0.10) QUES (0.10)	REAL (0.20) AS (0.11)	Values are referred to by labels, and sometimes pose questions, but are followed by insights or assumptions.
DEP	LL (0.12) TACT (0.11) ITP (0.09)	INC (0.17) AS (0.13)	Dependencies tend to be hard to grasp, but speculating is often applied.

indicate that insights follow assumptions or aggregation of information, however, are often followed by more confusion. Vice versa, contradictions follow realizations, but are often followed by further incomprehension.

When analyzing the relative prevalence of comprehension indicators we could observe recurring sequences, *i.e.*, comprehension patterns. Analyzing these patterns over all participants, we found that some patterns occur individually while others have been used by multiple participants. These common comprehension patterns hint at comprehension strategies. Table III shows a subset of comprehension patterns with their number of occurrences and number of participants that used it.

With these seven comprehension patterns, we can answer our research question: We observe that, for example, while consolidations (CONS) and realizations (REAL) often follow assumptions (AS), contradictions (CONTR) usually follow realizations (REAL). Furthermore, we observe that users often make an assumption (AS) about a label (LL) which can then help to make a realization (REAL) about the spreadsheet.

TABLE III: Number of occurrences (Occ.) of comprehension patterns and number of participants (Part.).

Pattern	Name	Occ.	Part.
AS ⇒ CONS	Assumptions	27	6
AS ⇒ REAL	Assumptions	22	6
REAL ⇒ CONTR	Contradiction	16	5
QUES, {LL,FF,DEP} ⇒ INC	Questioning	13	7
{CONS, ITP} ⇒ REAL	Reasoning	10	5
AS, LL ⇒ REAL	Conjecture	9	4
FF, CONS ⇒ REAL	Formula Abstraction	8	3

According to the number of occurrences and the number of participants, we found these seven comprehension patterns as potential comprehension strategies: For example, assumptions are often involved in the comprehension strategies and realizations often happen directly after. To avoid that such realizations are followed by contradictions, we recommend to provide appropriate tooling that supports the users with their comprehension strategies. Furthermore, questioning either formulas, the value of labels or dependencies often happens before incomprehension, which can indicate that users need support when analyzing the spreadsheets. We leave a more detailed investigation of these comprehension strategies to future work.

IV. Conclusions

In this paper, we conducted think-aloud study with 8 participants who were tasked to understand an unfamiliar spreadsheet. Analyzing the 8 transcripts from the study, we manually derived 16 comprehension indicators. With these indicators, we set out to identify frequent comprehension patterns that hint at comprehension strategies. We found, for example, that insights often follow assumptions and that they are often followed by confusion. With these results, we have taken the first step towards a holistic understanding of spreadsheet comprehension.

A. Threats to Validity

We selected participants and spreadsheets for this study which can bias our results. We mitigated this threat by selecting a diverse range of participants and various spreadsheets. Furthermore, we considered the transcripts from only 8 participants in this preliminary study. We plan to increase the validity by conducting the experiment with additional participants. Additionally, if participants avoided to comment during the experiment, we may have missed their thoughts. We mitigated this threat by encouraging the participants to actively share their thoughts.

B. Future Work

We plan to expand the study to the other two spreadsheets and 16 participants to obtain more generalizable results, and to triangulate our results with the results of Srinivasa *et al.* [13]. Furthermore, we plan to study the consequences of particular comprehension patterns, in particular concerning the supportive tooling when working with spreadsheets. Finally, we plan to detail the investigation of the found comprehension strategies.

Work partially funded through the Austrian Science Fund (FWF) project Interactive Spreadsheet Debugging (P 32445).

References

[1] R. T. Mittermeir, "Software evolution: A distant perspective," in *Proceedings of the International Workshop on Principles of Software Evolution*. IEEE, 2003, pp. 105–113.

[2] A. Kohlhase, M. Kohlhase, and A. Guseva, "Context in spreadsheet comprehension," in *Proceedings of the Workshop on Software Engineering Methods in Spreadsheets*, vol. 1355. CEUR-WS.org, 2015, pp. 21–27.

[3] F. Hermans, M. Pinzger, and A. van Deursen, "Measuring spreadsheet formula understandability," *CoRR*, vol. abs/1209.3517, 2012. [Online]. Available: http://arxiv.org/abs/1209.3517

[4] T. Isakowitz, S. Schocken, and H. C. Lucas, Jr., "Toward a logical/physical theory of spreadsheet modeling," *ACM Transactions on Information Systems*, vol. 13, pp. 1–37, January 1995.

[5] M. Tukiainen, "Comparing two spreadsheet calculation paradigms: An empirical study with novice users," *Interacting with Computers*, vol. 13, pp. 427 – 446, 2001.

[6] C. Kelleher and R. Pausch, "Lowering the barriers to programming: A taxonomy of programming environmtnets and languages for novice programmers," *ACM Computing Surveys*, vol. 37, no. 2, pp. 83–137, 2005.

[7] S. Vemuri, S. Sengupta, and J. S. Davis, "Data dependency diagrams for spreadsheet applications," in *Proceedings of the Annual Southeast Regional Conference*. ACM, 1992, pp. 467–470.

[8] D. G. Hendry and T. R. G. Green, "Creating, comprehending and explaining spreadsheets: a cognitive interpretation of what discretionary users think of the spreadsheet model," *International Journal of Human-Computer Studies*, vol. 40, no. 6, pp. 1033 – 1065, 1994.

[9] M. Burnett, C. Cook, O. Pendse, G. Rothermel, J. Summet, and C. Wallace, "End-user software engineering with assertions in the spreadsheet paradigm," in *Proceedings of the International Conference on Software Engineering*. IEEE, 2003, p. 93–103.

[10] B. Bishop and K. McDaid, "Unobtrusive data acquisition for spreadsheet research," in *Proceedings of the Symposium on Visual Languages and Human-Centric Computing*. IEEE, 2008, pp. 139–142.

[11] ——, "Spreadsheet debugging behaviour of expert and novice end users," in *Proceedings of the International Workshop on End-user Software Engineering*. ACM, 2008, pp. 56–60.

[12] F. Hermans, M. Pinzger, and A. van Deursen, "Breviz: Visualizing spreadsheets using dataflow diagrams," *CoRR*, vol. abs/1111.6895, 2011. [Online]. Available: http://arxiv.org/abs/1111.6895

[13] S. Srinivasa Ragavan, A. Sarkar, and A. Gordon, "Spreadsheet comprehension: Guesswork, giving up and going back to the author," in *Proceedings of the International Conference on Human Factors in Computing Systems*. ACM, 2021, p. to appear.

[14] F. Hermans and E. Murphy-Hill, "Enron's spreadsheets and related emails: A dataset and analysis," in *Proceedings of the International Conference on Software Engineering*, vol. 2. IEEE, 2015, pp. 7–16.

[15] D.-Y. Chiu, Y.-H. Wu, and A. L. Chen, "An efficient algorithm for mining frequent sequences by a new strategy without support counting," in *Proceedings of the International Conference on Data Engineering*. IEEE, 2004, pp. 375–386.