5th Human Factors and Ergonomics Malaysia Conference & ErgoLympic 2023, 15-18 August, 2023. Langkawi, Malaysia.

The Usability of Voice User Interface for Personal OKRs Setting

Sakol Teeravarunyou[1*], Kochahem Kamolwit[1]

[1]King Mongkut's University of Technology Thonburi, Bangkok, Thailand
sakol.tee@kmutt.ac.th, kochahem.kam@kmutt.ac.th

Abstract. The process of defining objectives and key results (OKRs) can be a challenging task for users, particularly for millennials who may struggle to set clear goals. Despite the seemingly straightforward definition of OKRs, many individuals find it difficult to establish their objectives and key results. In response, this study aims to develop a Voice User Interface (VUI) application that assists users in setting their OKRs, while also examining the ease of use of the microcopy, utterance, and visual display. The study involved 25 participants who used the application to set and track their goals over a period of one month. Based on the results of the System Usability Scales (SUS), the application was found to be effective and user-friendly. Participants provided feedback on how to improve the product, such as making the system more conversational, using a human-like tone, condensing utterances, and shortening microcopy to better match users' language.

Keywords: Voice user interface, OKRs, Usability, Utterance, Microcopy.

1 Introduction

The changing nature of work in the modern era has resulted in an increasing demand for new skills among the workforces. This has created a potential mismatch between the skills that are available and those that are needed by employers. In response to this challenge, members of millennials are turning to self-directed learning as a means of acquiring the necessary skills to remain competitive in the job market. Self-directed learning is an approach in which individuals take responsibility for identifying their own learning needs, setting learning goals, finding the necessary resources, and evaluating their knowledge. This approach can involve studying online courses, reading books, listening to podcasts, or engaging in on-the-job training. To support this self-directed learning effort, a framework and tools are required to help individuals succeed. In this research, the goal-setting theory and management by objectives are used to motivate participants in their self-directed learning efforts. One such success framework is the Objectives and Key Results (OKR) approach. This approach has been used by organizations such as Intel and Google to measure and track performance outcomes [1].

In formal education settings, such as schools, OKRs are also used to set and measure performance goals. OKRs software, such as Profit.co [2] and Weekdone [3], are available to assist individuals in monitoring and evaluating their progress. However, the existing software lacks the inclusion of habit-forming strategies in personal OKRs. While many of these software solutions cater to individual OKRs, there is currently a gap in the market for personal OKRs software that incorporates habit formation techniques. Individual OKRs primarily focus on the goals and objectives established by an individual within a team or organization. These objectives align with the broader aims of the team or company, contributing to overall success. On the other hand, personal OKRs extend beyond the professional sphere, encompassing an individual's personal life, interests, and aspirations. These OKRs represent self-defined goals set by individuals across various aspects of their lives, such as health, personal growth, relationships, or hobbies.

Therefore, the purpose of this research is to develop a Personal OKRs software specifically designed to help members of millennials set career development goals and personal OKRs. The software has two objectives: to elicit goals and key results that users want to achieve and to facilitate the habit tracking of self-directed learning. While the definition of OKRs may seem straightforward, the challenge lies in defining the right key results. Previous research has shown that the SMART goals framework can be used with key results of OKRs [4]. The SMART framework emphasizes specific, measurable, attainable, results-based, and time-bound [5].

To further enhance the user experience, a voice user interface (VUI) is used to elicit goals and key results. VUI provides a more natural and intuitive means of interaction, which can increase task completion rates and reduce time and effort [6]. To ensure clear communication and minimize confusion among users, microcopy is used to explain and guide users through the OKR setting process. The research focuses on the effect of VUI on input data and how it can be used to facilitate the OKR tracking process. A well-designed VUI can guide users in setting and tracking their OKRs, while providing real-time feedback to aid in habit formation [7]. The research objective is to design a VUI that is user-friendly and effective in terms of OKR settings, and to examine the components of VUI such as microcopy, audio, and visual display in terms of usability.

2 Method

2.1 Participants

The study sample comprised individuals belonging to the millennial's cohort, ranging from 21 to 37 years of age, who possessed access to and demonstrated proficiency in using digital technologies, such as the internet, mobile phones, and computers. The participants were employed in various fields, including working for companies, freelancing, or running their own businesses. The study recruited a total of twenty-four participants, including nineteen females and six males. In appreciation of their participation, each participant received a $94 incentive. The research protocol received ethical approval from the Institutional Review Board at the University (KMUTT-IRB-2022/07011/219).

2.2 Instrument

The research methodology employed in this study was the Wizard of Oz method [8–10]. The prototyping application was designed to simulate smart display devices, akin to Google Nest Hub and Amazon Echo Show. Participants utilized their personal computers while two test monitors oversaw the study remotely. One test monitor was responsible for selecting the utterance playlists of the voice assistant, while the other test monitor was tasked with inputting and tracking the subject's conversation data. Behind the scenes, the test monitor calculated the key results levels.

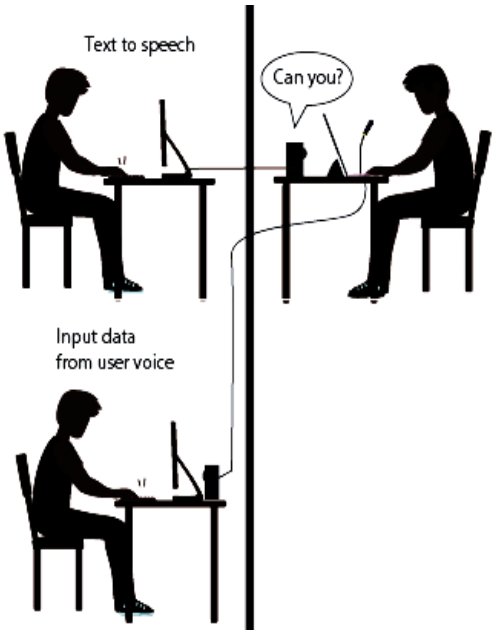

Fig. 1. Wizard of Oz method with two monitors (left) while subject (right) with remote access from internet.

The experiment in this study was conducted via the Zoom meeting platform (refer to Fig. 1). Participants were informed that they were situated in their own homes and were prompted to interact with the smart display through a dual modality of audio and graphical user interfaces. Both utterances, or audio conversation, and microcopy, or label instructions, were implemented by researchers. In cases where participants were unable to understand the utterance, they could rely on the accompanying microcopy for guidance.

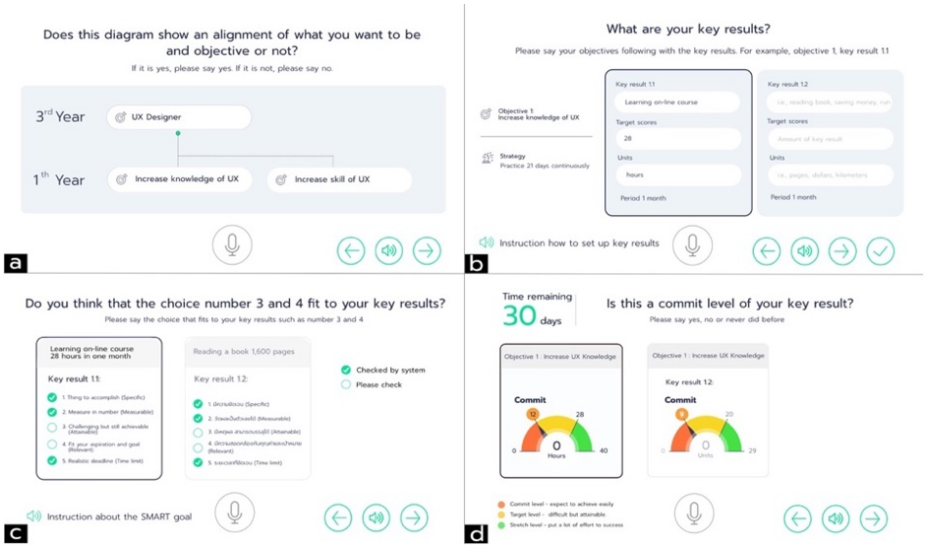

Fig. 2. Screenshot of VUI display a) goal alignment, b) key results, c) SMART framework and d) checking of commit level.

Prior to the experimentation, researchers provided participants with instructions on Objectives and Key Results (OKRs) one week in advance. These instructions included technical terms associated with OKRs and examples of key results. The application used in the experiment is displayed in Figure 2. Upon initiation, the system prompted participants to provide responses, and subsequently generated a goal alignment in the form of a tree diagram (refer to Figure 2a). Participants were able to view the placeholder for key results as an example. After indicating their respective key results, including the amount and unit, the system automatically populated the form (as shown in Figure 2b). In addition, the computer aided participants through the provision of a checklist interface and requested participants to confirm whether their established goals adhered to the SMART framework. Several options were already pre-checked by the system (as illustrated in Figure 2c). The gauge utilized indicated the commitment level of the key result. Key result scoring criteria consisted of three levels: 1) Stretch - representing the most optimal outcome that could be obtained (with only a 10% likelihood of achievement), 2) Target - reflecting a challenging but attainable outcome (with a 50% likelihood of success), and 3) Commit - indicating an anticipated outcome achievable with minimal effort (with a 90% likelihood of success). Participants set their target value, and the system automatically calculated both the commit and stretch values. Participants could indicate 'yes' if they felt their committed values were reasonable or indicate 'never did before' if they lacked experience in undertaking the key result (as shown in Figure 2d).

2.3 Procedure

The present study involved the implementation of Objectives and Key Results (OKRs) by the participants, encompassing the formulation of goals, identification of key results, and selection of habit strategies. Following this, the participants were administered the System Usability Scales (SUS) questionnaire to evaluate the usability of the system. The participants were required to establish a goal and determine 1-3 key results, in addition to selecting from a pool of three habit strategies, namely 1% improvement per day, 26 days of continuous practice, and time blocking.

2.4 Analysis

The present investigation employed the System Usability Scale (SUS) as a usability metric, which is widely applicable across an array of digital systems and applications, ranging from mobile applications and laptops to machinery [11, 12]. The SUS is a standardized tool comprising ten questions with a five-point rating scale that ranges from strongly disagree to strongly agree (1-5). It is designed to assess user perception and satisfaction with a given system's usability and has been extensively utilized in prior research endeavors [12, 13]. The calculation of the score is.

$$SUS\ score = 2.5(x + y) \quad (1)$$

where:

x = *Sum of the points for all odd-numbered question – 5*

y = *25 – Sum of the points for all even-numbered questions*

In this study, Equation (1) was employed to calculate a total score of 100, where each of the ten questions in the System Usability Scale (SUS) questionnaire contributed 10 points. The odd-numbered questions were positively toned, while the even-numbered questions were negatively toned. The participant's raw scores for each question were transformed into a new number, which were then summed and multiplied by 2.5, to obtain a score on a scale of 0-100. A SUS score greater than 68 was considered above average. Following the completion of the usability evaluation, participants were requested to rate their satisfaction with the overall system.

3. Results

Table 1. The mean score of SUS questions.

SUS questions	Mean (SD)
1. I think that I would like to use this system frequently.	3.66(.833)
2. I found the system unnecessarily complex.	2.84(21)
3. I thought the system was easy to use.	3.9(.18)
4. I think that I would need the support of a technical person to be able to use this system.	3.05(.31)
5. I found the various functions in this system were well integrated.	4.43(.13)
6. I thought there was too much inconsistency in this system.	1.7(.18)
7. I would imagine that most people would learn to use this system very quickly.	3.69(.19)
8. I found the system very cumbersome to use.	2.14(.24)
9. I felt very confident using the system.	4.06(.2)
10. I needed to learn a lot of things before I could get going with this system.	2.28(.26)

In this study, the summative scores for the positively toned and negatively toned questions in the System Usability Scale (SUS) questionnaire were calculated as 19.77 and 12.07, respectively. The total SUS score was then determined to be 69, which is considered good and acceptable, as depicted in Table 1. The scores obtained from the ten individual questions are presented below:

Question 1: I think that I would like to use this system frequently.

In the present investigation, participants were requested to rate their usage of the system in response to the first question (M=3.66, SD=.833). Among the twenty-five participants, seven expressed their intention to use the system in the future, citing that it aided them in achieving their self-improvement goals and providing clear guidance. However, an equal number of participants opined that they required additional guidance and found certain features of the system difficult to comprehend. The remaining participants stated that they may set goals less frequently, over a quarter, and were unsure about their future usage of the system.

Question 2: I found the system unnecessarily complex.

In response to the second question, the participants rated the system with a mean score of 2.84, indicating that they perceived the system to be complex. The lack of prior experience in setting objectives and key results (OKRs) posed a challenge for many participants, resulting in difficulty in setting up their key results accurately. Participants expressed their concerns regarding the technical language employed in OKRs, such as the terminology used for key results and goals. While nine participants believed that the

system employed overly formal language and recommended a more streamlined process for setting OKRs. Ten participants found the flow of the software acceptable, with an easy-to-understand visual interface that reduced perceived complexity.

Question 3: I thought the system was easy to use.

In the present study, a significant proportion of the participants rated the VUI system as being easy to use (*M*=3.9, *SD*=.18*, MO*=4.33). The majority of the fourteen participants provided feedback indicating that the speech-to-text technology employed in the system made it easy to use, as it facilitated a conversational interaction with the smart device. However, these participants recommended supplementing the voice commands with text descriptions and illustrations, as it may be challenging to remember all commands. Overall, these measures could improve communication and comprehension.

Question 4: I think that I would need the support of a technical person to be able to use this system.

In the present study, the participants provided a moderate request for technical support (*M*=3.05, *SD*=.31, *MO*=5). The main difficulty experienced by the participants was related to setting a goal and comprehending the technical terminology used in the system. Nine of the participants expressed their desire for technical support, emphasizing the importance of receiving explanations from technical supporters regarding the meaning of the technical terms used. They suggested that tooltips or other informative features could help with this issue. Furthermore, despite the presence of a video clip that explains the system before its use, three participants expressed a preference for a tutorial or onboarding process at the beginning of the app use. While some participants found the system to be clearly explained, the concept of OKRs setting required a level of understanding and experience, particularly regarding the unit of key results and committed levels. As a result, technical support or a manual remained necessary.

Question 5: I found the various functions in this system were well integrated.

Based on the participants' responses, it was found that they generally agreed that the functions and features of the system were well integrated (*M*=4.43, *SD*=.13, *MO*=5). Specifically, many participants noted that the sequence, flow, and navigation of the system were seamless, and the integration of habit strategies, key results, and SMART goals was well executed. They also appreciated the summary of all activities, which allowed for a clear understanding of the entire process. However, three participants expressed a desire for additional features, such as the ability to generate reports of their OKRs setting and tracking progress.

Question 6: I thought there was too much inconsistency in this system.

Participants rated the level of inconsistency in the system as relatively low (*M*=1.7, *SD*=.18). Only a participant identified a minor inconsistency related to the layout of image, whereas others pointed out the delayed response time of the voice assistant. Participants acknowledged that the application maintained a consistent mood and tone throughout the user journey, indicating that the system achieved a high level of internal

consistency. However, they expressed concern about the latency issues when processing natural language.

Question 7: I would imagine that most people would learn to use this system very quickly.

Regarding the learning ability of the system, the participants' assessment indicates that individuals can learn to use the system quickly (*M*=3.69, *SD*=.19). However, it should be noted that prior knowledge of OKRs is essential, and therefore, an instructional manual or onboarding process is crucial before using the application. The participants also recommended that the microcopy within the app be concise and straightforward for ease of understanding. They expressed concern that older generation without much experience in VUI may have difficulty using the system.

Question 8: I found the system very cumbersome to use.

The results indicate that participants did not find the system cumbersome to use (*M*=2.14, *SD*=.24). One participant expressed that the voice assistant was helpful in guiding their thought process and did not find the system cumbersome. However, three participants noted that the voice assistant sounded like a robot and required careful attention. Additionally, one participant reported feeling uncomfortable if he was unable to respond to the voice assistant's questions.

Question 9: I felt very confident using the system.

According to the results of the study, participants reported feeling confident while using the system (*M*=4.06, *SD*=.2). The participants appreciated the system's ability to facilitate form completion, provide a recheck of input, and confirmation of responses. Specifically, some participants noted their preference for the checklist of SMART goals, which helped to clarify the goal and provide greater confidence in the planning process. Overall, the system was deemed to have clear and well-defined steps and sequences, leading some participants to express confidence in their ability to work with the system independently.

Question 10: I needed to learn a lot of things before I could get going with this system.

The level of learning required before using the system was deemed relatively low (*M*=2.28, *SD*=.26). Five participants suggested that while extensive learning was not necessary, a better understanding of the OKRs framework and taking time to reflect on one's goals was helpful.

4. Discussion

Communication presents a significant challenge in the use of the OKRs setting, where technical terms were not well understood by participants. To optimize communication, the utterance of the voice user interface (VUI) should be concise and straightforward, as participants were disinclined to read or listen to long explanations more than once.

This issue is consistent with previous research on accessibility in mobile applications [13], where the primary problem encountered by users related to alternative vocabulary suggestions (37.9%). One way to mitigate this is through modifications to the textual content, including the replacement, removal, or addition of terms to make them more accessible to users. Additionally, users have expressed a preference to use language familiar to them rather than technical terms like OKRs. Participants with prior experience with OKRs training were found to be less affected by these challenges. The length of commands in VUIs was also problematic as participants struggled to memorize lengthy sentences, which highlights the limitations of human memory. Previous research has suggested the need to reduce the cognitive load of participants when formulating commands. Participants also expressed dissatisfaction with the efficacy of some utterances, which often required considerable thought to interpret. To address this, most participants preferred a multimodal platform that combined audio with visual aids such as graphics and microcopy [14], although some participants preferred audio alone.

Conciseness is crucial when considering the length of microcopy, which should reflect the user's preferred style. This is because style is an essential aspect of an individual's self-presentation and social identity [15]. For instance, cookie consent notices can adopt different styles ranging from casual, informal, and relaxed to more formal and serious tones. When considering OKRs, it is essential to ensure that the style matches the preferences of the millennial’s demographic, who tend to favor a more casual tone and have a faster understanding of information. Some participants reported difficulties with long and formal utterances, while others had trouble relating to the technical terminology used. Therefore, it may be necessary to modify the vocabulary and style of OKRs to align with users' mental models. Additionally, participants suggested that the sound of VUIs should be more human-like and less robotic. Previous research indicates that users struggle to interpret voice without additional information such as tone, volume, intonation, and speech rate [16]. This highlights the importance of considering the emotive voice of agents. The sound of the voice assistant also needs to be customized for specific cultural preferences. For example, in this experiment, the Thai language voice assistant sounded different from its English counterpart, which may require modification to ensure that the essence and tone of VUI match the user's native language. One participant suggested that the sound should be akin to a voiceover on a live stream game.

Voice user interface (VUI) offers several advantages over traditional user input methods. For instance, VUI allows participants to interact with the system without the need for manual keyboard input. Furthermore, the system can detect the level of difficulty of questions by measuring the delay in the participants' responses. VUI also facilitates skill elicitation, as the system calculates the key result level of the target based on the commit level. When designing the checklist interface, it may be beneficial to adopt a pilot checklist format. Instead of relying on participants to read through the checklists, the VUI system can read the checklists for them and ask for their responses, which can simplify the process and enhance comprehension.

The present system necessitates technical support to facilitate the comprehension of the OKRs concept, which proved to be a challenge for the participants in setting their

units of key results. For instance, the quantification of reading a book could entail various units of measurement, such as the number of characters, pages, or duration time, each of which can potentially affect the tracking of OKRs. If the unit of measurement was overly detailed, it would cause inconvenience for the participants in collecting data, whereas a rough unit would lead to inaccurate data that is automatically linked to many data collection apps. The level of commitment also emerged as an issue that impacted the participants' attainment of OKRs, as it required prior experience with the practice. In the experiment, for instance, one participant lacked experience in writing travel blogs, and thus could not estimate the appropriate committed level. Despite the tutorial provided before using the application, participants still required practice in writing key results. An alternative solution could be to develop a standardized key result that participants can choose from.

To enhance future development, it is recommended that the VUI incorporate personalized features and styles that cater to specific user groups. The system should maintain an archive of key results that can serve as a valuable resource for novice users. In addition, integrating data from other applications and smart devices into the system can facilitate a more comprehensive understanding of users' situations. Ultimately, the system holds the potential to support members of millennials in achieving their future goals and aspirations.

5. Conclusion

The Wizard of Oz method proves to be effective in simulating the interaction between participants and the test monitor. Many participants perceived that they were engaged in a conversation with an intelligent machine. The System Usability Scale (SUS) score of the VUI was calculated to be 69, which, from a holistic perspective, indicates acceptable performance but still leaves room for improvement. The first area for improvement pertains to interface design, specifically with respect to employing concise utterances and microcopy, adopting a human tone of voice, and streamlining the checklist. The second area for improvement involves domain-specific knowledge, including the ability to comprehend OKR terminology, write correct key results, and interpret key result levels. The VUI serves to assist users in eliciting their goals and framing their objectives and key results. In future development, the VUI could be utilized to track users' habits and devise habit-forming strategies, thereby enhancing the application's effectiveness in daily self-improvement tasks such as learning and habit-forming.

Acknowledgments. This research project is supported by Thailand Science Research and Innovation (TSRI) - Basic Research Fund: Fiscal year 2022 under project number FRB650048/0164

References

1. Duke, B.N.: Analyzing the Impact of OKR Goal Setting with Teachers on student and teacher performance outcomes., (2021)
2. Profit.co: OKR Software that helps you execute your strategy, https://www.profit.co, (2023). Accessed 18 Feb 2023
3. Kaljundi, J.: OKR Software for Real Company Results, https://weekdone.com, (2023). Accessed 18 Feb 2023
4. III, B.D.: Analyzing the Impact of OKR Goal Setting with Teachers on Student and Teacher Performance Outcomes. (2021).
5. Conzemius, A., O'Neill, J.: The power of SMART goals: Using goals to improve student learning. Solution Tree Press (2009).
6. Pyae, A., Joelsson, T.N.: Investigating the Usability and User Experiences of Voice User Interface: A Case of Google Home Smart Speaker. Presented at the (2018).
7. Stawarz, K., Cox, A.L., Blandford, A.: Beyond Self-Tracking and Reminders: Designing Smartphone Apps That Support Habit Formation. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing. New York, USA. (2015). https://doi.org/10.1145/2702123.2702230
8. Berdasco, A., López, G., Diaz, I., Quesada, L., Guerrero, L.A.: User Experience Comparison of Intelligent Personal Assistants: Alexa, Google Assistant, Siri and Cortana. Proceedings 2019, Vol. 31, Page 51. 31, 51 (2019). https://doi.org/10.3390/PROCEEDINGS2019031051
9. Park, S., Lim, Y.K.: Investigating User Expectations on the Roles of Family-shared AI Speakers. Conference on Human Factors in Computing Systems - Proceedings. (2020). https://doi.org/10.1145/3313831.3376450
10. Choi, S., Kim, Y., Son, J., Cho, J., Han, E.: Poster: Designing conversational voice user interface for improving intimacy of shared invehicle. UbiComp/ISWC 2018 - Adjunct Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2018 ACM International Symposium on Wearable Computers. 33–37 (2018). https://doi.org/10.1145/3267305.3267638
11. Bangor, A., Kortum, P.T., Miller, J.T.: An Empirical Evaluation of the System Usability Scale. https://doi.org/10.1080/10447310802205776. 24, 574–594 (2008). https://doi.org/10.1080/10447310802205776
12. Kortum, P.T., Bangor, A.: Usability Ratings for Everyday Products Measured With the System Usability Scale, 67–76 (2013). http://dx.doi.org/10.1080/10447318.2012.681221
13. Tymoshchuk, O., Oliveira, E., Branco, A., Carvalho, D., Antunes, M.J., Pedro, L., Almeida, A.M., Ramos, F.: Accessibility and microcopy remote testing of mobile applications: The case of the CeNTER platform. Iberian Conference on Information Systems and Technologies, CISTI. (2021). https://doi.org/10.23919/CISTI52073.2021.9476413
14. Tangmanee, K., Teeravarunyou, S., Buaban, N.: Voice User Interface (VUI): A Review of Present and Potential Voice Assistant (VA) applications. Thai Journal of Ergonomics. 3, 32–44 (2020).
15. Portmann, L.: Crafting an audience: UX writing, user stylization, and the symbolic violence of little texts. Discourse, Context and Media. 48, (2022). https://doi.org/10.1016/J.DCM.2022.100622
16. Wang, Z.: Future Challenges in the Next Generation of Voice User Interface. In: 2020 International Conference on Computing and Data Science (CDS). pp. 191–193. , CA, USA (2020).