Involving Senior Workers in Crowdsourced Proofreading

Toshinari Itoko[1], Shoma Arita[2], Masatomo Kobayashi[1], and Hironobu Takagi[1]

[1] IBM Research – Tokyo, 5-6-52 Toyosu, Koto, Tokyo 135-8511, Japan
{itoko,mstm,takagih}@jp.ibm.com
[2] The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-8656, Japan
arita@cyber.t.u-tokyo.ac.jp

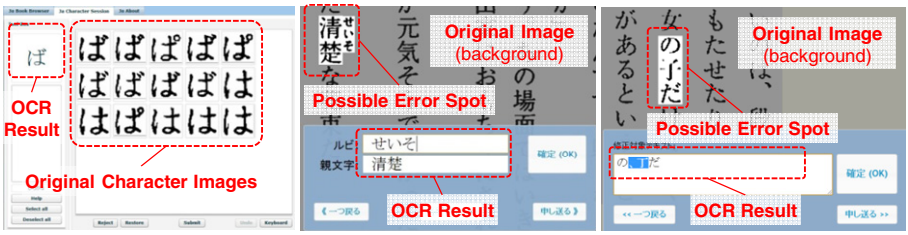
Abstract. Seniors have a wealth of knowledge and free time, so they are a promising workforce for crowdsourced tasks. Currently senior workers are hardly involved in real applications. We have started an experimental project that crowdsources proofreading micro-tasks to volunteer workers to efficiently produce accessible digital books. By design, the majority of the workers in this project are senior citizens. In this paper, we report the findings of our experiment in which we tested four working hypotheses about the behavioral characteristics of senior workers. We also discuss skill management to improve task performance and motivation encouragement for long-term involvement of senior workers.

Keywords: Senior Workforce, Elderly, Ageing, Micro-tasks, Crowdsourcing, Gamification, Accessibility.

1 Introduction

Crowdsourcing is recognized as a powerful tool for outsourcing manual tasks and is widely utilized in real applications. The tasks offered by typical crowdsourcing services are not demanding as regards the workers' physical locations or time, so they are highly suitable for a large number of non-fulltime workers. Seniors are an especially promising workforce for such tasks, especially in Japan where the population is aging rapidly [1]. However, there are no well-established methods for effectively involving senior workers in crowdsourced work. Our findings will help accelerate the development of methods to support crowdsourced applications with. In October 2013, in collaboration with the Japan Braille Library, we launched an experimental crowdsourcing project to convert printed books into an accessible digital text format, DAISY (Digital Accessible Information SYstem) [2], using the micro-tasking model proposed in [3]. As of January 2014, 178 participants including 83 seniors (age 60+) have registered and more than 1,200 hours of work has been completed. Since the crowdsourcing of proofreading is a well-known approach (e.g., [4][5][6]), our research focus was to develop methods to improve senior workers' long-term performance in crowdsourcing. These methods must be based on a deep understanding of their behaviors in practical applications. This paper is organized as follows. In Section 2, we introduce the implementation of our system and the experimental hypotheses.

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2014, Part III, LNCS 8515, pp. 106–117, 2014.



(a) Character (b) Ruby (c) Phrase

Fig. 1. Three types of crowdsourced proofreading interfaces

In Section 3, we report the results and notable findings. In Section 4, we discuss how to manage skills and motivate senior workers doing crowdsourced tasks.

2 Implementation and Hypotheses

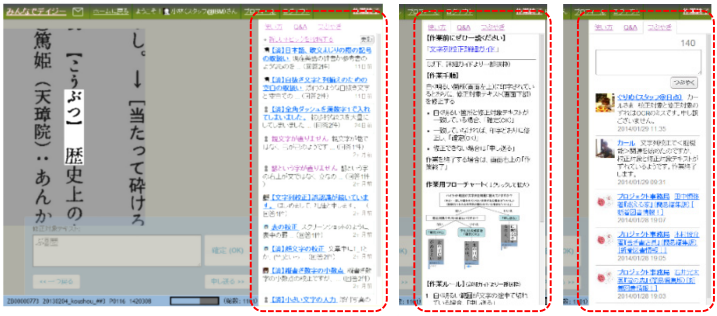
Our experiment is designed to understand behavioral features of senior workers from two main perspectives: proofreading operations and mechanisms for encouraging motivation. Starting with the system design proposed in [3], we introduce additional implementation considerations and experimental hypotheses.

Total Involvement. We arranged the experimental website to improve the involvement of senior workers as shown later in this section. Given that senior citizens tend to have a desire to contribute to their society [7], seniors are expected to work more if the technical and motivational barriers are eliminated. Thus we start with this hypothesis:

(H1). Seniors will do more work than young workers.

2.1 Crowdsourced Proofreading Interfaces

Our system decomposes the proofreading process into three types of sub-tasks and provides a specialized view for each of them: Character, Ruby, and Phrase (Fig. 1). This is based on the observation that there are three kinds of typical OCR (Optical Character Recognition) errors in Japanese text: (a) repeated errors involving a specific character (i.e., a specific character tends to cause similar OCR errors throughout a book), (b) errors in ruby (a pair consisting of one or more Chinese-derived characters and a pronunciation gloss (displayed nearby in a smaller (ruby) font)), and (c) letter separation errors in phrases (as in English when “m” is read as “r” plus “n”). Since the Character view uses an interface from CONCERT [5], it has some English labels, which were not redesigned for the Japanese users. In comparison, the Ruby and Phrase views have relatively large buttons with Japanese labels. See [3] for details of the design of each view.



(a) Q&A (b) Usage (c) Tweets

Fig. 2. Screenshots of the sidebar of the proofreading interfaces

Differences in Competence. Several articles indicate that senior workers have more linguistic knowledge but weaker ICT skills and visual attention than younger workers [8][9]. We believe that Character requires more ICT skills and visual attention while Ruby and Phrase call for more linguistic knowledge, which leads to our second hypothesis:

(H2). Seniors will be relatively good at the Ruby and Phrase tasks while young workers are better at Character.

2.2 Mechanisms for Encouraging Motivation

With all three of our proofreading views, workers can start or leave the work whenever they want to. There are no quotas or scheduled hours. This makes participation easier, especially for people who have little time for volunteer work. However, this also means that they can always leave the project, and so we added additional mechanisms to encourage their long-term involvement. We used two approaches to achieve this goal: removing barriers to continuing participation and providing incentives for active participation. For barrier removal, our system has question-answer support, since a leading cause of quitting is when a worker cannot complete a difficult task. For incentive, a gamification mechanism provides workers with a variety of types of feedback in response to their contributions.

Question-Answer Support. Previous studies have indicated that senior citizens using ICT have stronger needs for support from other people compared to younger people [10][11][12], we carefully integrated a question-answer (Q&A) forum into the proofreading interfaces as illustrated in Fig. 2-a. The list of the latest questions, a link to the full list of questions, and a link to post a new question are always visible at the right side of the proofreading interfaces. This allows proofreaders to easily access the question-answer features whenever they need to during proofreading tasks. The Q&A forum can transfer knowledge about the language itself, the usage of the system, and the rules of proofreading, coming from the participants who know the answers to the

participants who need to know. We anticipated that senior participants would post relatively more questions in the forum since they prefer guided support from other people and tend to dislike trial-and-error approaches [10]. In the sidebar, there are two other tabs. "Usage" shows brief instructions for the ongoing proofreading task with a link to the full instruction manual in PDF format (Fig. 2-b). Again, research shows senior citizens prefer to learn from manuals almost as much as they like support from other people, but they prefer to print documents rather than read them on screen. The "Tweets" tab offers a Twitter-like chat interface (Fig. 2-c) intended for frequent, informal communications among participants that could lead to a stronger sense of collaboration. Regarding the Q&A forum, we have a third hypothesis based on the previous research:

(H3). Seniors will ask more questions than the young workers.

Gamification. Rankings and badges are popular methods to motivate active participation in various types of online social systems. It is known that providing participants with feedback about their contributions can improve their motivation (e.g., [13]). Since crowdsourcing allows workers to freely start and leave their work whenever they want, motivation is needed for sustainable involvement. Thus we added gamification features to our system. They consist of (a) a ranking based on the number of completed tasks during the last 30 days, (b) badges based on completed tasks, (c) the accumulated numbers of completed tasks of each task type, (d) the worker's personal contribution to the last book the worker contributed to, as a percentage of all of the work done so far on that book, and (e) the number of books that the worker has contributed to relative to the total number of books in this project (Fig. 3). Note that (a) is a kind of competition between each worker and the other workers while (b) and (c) measure the worker's own efforts. The scores of (d) and (e) give a larger perspective on the worker's contribution. Since previous studies (e.g., [14]) showed that gamification mechanisms could benefit senior citizens as well as younger people, this leads to our fourth hypothesis:

(H4). Gamification features will more strongly encourage seniors than young workers.

2.3 Miscellaneous Design Considerations

Fig. 3 shows the portal page that the participants see after logging into the system. It consists of announcements to the participants at the top, large icons that link to the main features (e.g., proofreading or Q&A) in the middle, a gamification screen at the bottom, and links to download manuals and the chat interface on the right side. The portal page is designed not only for proofreaders but also for DAISY book users, i.e., people with print disabilities. The system provides them with forums to request new DAISY books, which can lead to new proofreading tasks. One of our goals is to keep this forum active, because a previous study reported that the lack of tasks can disrupt the momentum of participation in a micro-tasking system, and the lost

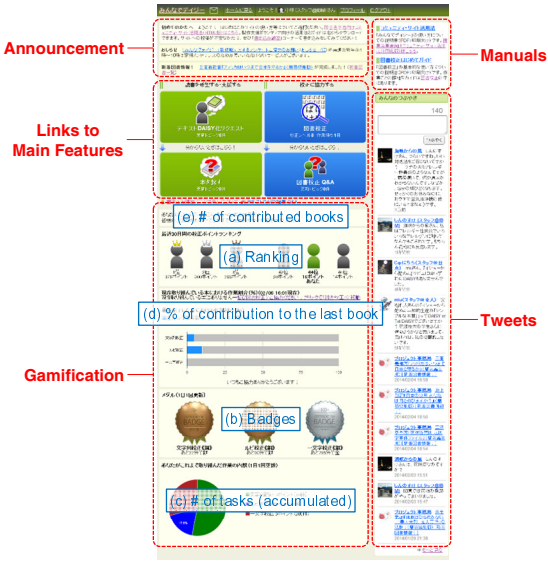


Fig. 3. A screen shot of the portal page with the gamification display

participants may never return [15]. Finally, the system also provides a sandbox forum to allow participants to practice using the forum interface.

The website design for senior citizens involved many accessibility-related considerations. For example, their declining sensory and physical abilities require larger text and buttons [16]. Meanwhile, they often use inexpensive devices with small screens (e.g., XGA displays). Senior citizens may have difficulties in scrolling the pages, which imposes a design constraint that all of the needed information should be visible on one screen. Thus there is a trade-off between the sizes of the content and the screen. We used an iterative design process based on user feedback with experts deciding on the size of each component and their layout. Also, we iteratively improved the accessibility for the DAISY book users, who usually use a screen reader to access websites.

3 Experimental Results

In this section, we describe the results of our experiment in terms of the four hypotheses (H1–4) from Section 2. The experiment started on October 15, 2013. We analyze the data from this date to January 26, 2014 in this paper. In that period, 178 volunteers registered and 112 of them were “active workers”, who did at least one task. They contributed more than 1,200 hours and proofread 136 books. The outcomes by types of tasks are summarized in Table 1.

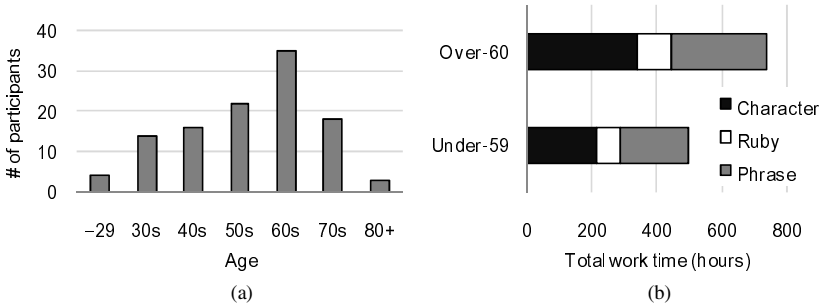


Fig. 4. (a) Age distribution of active workers and (b) Total work time (hours)

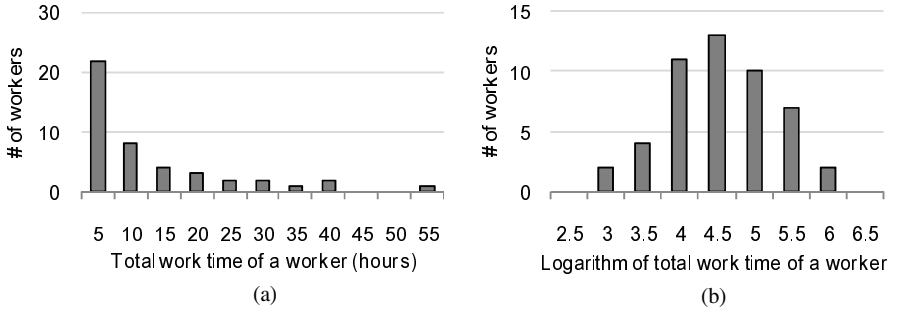


Fig. 5. Distribution of (a) the total work time (hours) and (b) the logarithm of (a)

Table 1. Proofreading Outcomes by Task Type

Task type	# completed	Total working time	# per hour
Character	279,390 tasks	553 hours	505 tasks
Ruby	77,930 tasks	180 hours	432 tasks
Phrase	137,070 tasks	500 hours	274 tasks

3.1 Total Involvement

The active workers had wide distribution of ages (Fig. 4-a). There were 56 active senior workers (Over-60) and 56 active younger workers (Under-59). The Over-60 group worked a total of 755 hours while the Under-59 group worked 514 hours (Fig. 4-b). The distribution of total work time for each active worker was not a normal distribution (Fig. 5-a). However, the logarithm of the total work time seemed to be normally distributed (Fig. 5-b), with statistical support from a Shapiro-Wilk normality test ($p > .05$). A Welch test showed that the logarithmic total work time for Over-60 was significantly longer than the Under-59 ($p < .05$). The average log values were 4.107 and 3.769 for Over-60 and Under-59, respectively.

(R1). H1 was supported. Over-60 worked significantly longer than Under-59.

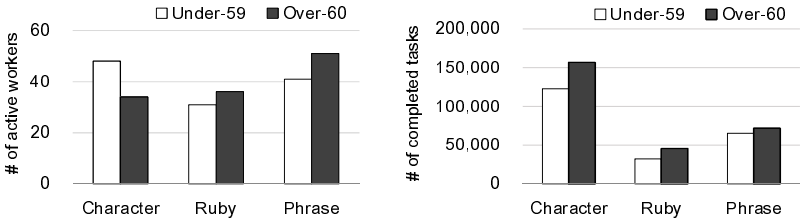


Fig. 6. (a) The number of active workers and (b) The total number of completed tasks for each task type

3.2 Differences in Competence

H2 seems to be supported as shown by the number of active workers for each task type in Fig. 6-a. In particular, for the participants older than 70, only 7 out of 21 tried to do Character tasks while almost all of them (20 out of 21) tried Phrase tasks. The χ^2 tests showed that Over-60 was significantly less active in Character tasks and more active in Phrase than Under-59 ($p < .05$). There was no significat difference in Ruby. H2 was not supported based on the total number of tasks completed for each task type in Fig. 6-b. The Over-60 group did more Character tasks than Under-59. A Wilcoxon rank sum test for each task type showed that there is no significant difference in the efficiency (number of completed tasks / work time) between Over-60 and Under-59 ($p > .05$).

(R2). H2 was partially supported. The Over-60 group was less likely to try Character tasks while they tried more Phrase tasks. However there was no significant difference in the efficiency metric between the Over-60 and Under-59 groups.

3.3 Question-Answer Support

During the experimental period, 60 questions were posted by proofreaders. The average numbers of questions for each active participant were 0.63 and 0.70 for Under-59 and Over-60, respectively. This result seems to confirm H3. However, the limited number of questions is too small to support a firm conclusion. Certain participants tend to post many questions, while 85% of the active participants did not post any. This result does not necessarily indicate that the Q&A forum is of no use for the majority of the participants. The data shows that 73% of the active participants accessed the Q&A forum and 59% read at least one Q&A thread. They may have gained knowledge from the Q&A forum even if they did not post any questions. In addition, several participants wrote question-like messages using the chat interface rather than the Q&A interface. This might indicate that if the system provides a quicker way to post questions, the participants would be encouraged to post more questions. Out of the total of 60 questions, 56 were answered by members of the library or development teams, only 1 was answered by other participants, but 3 received answers from both

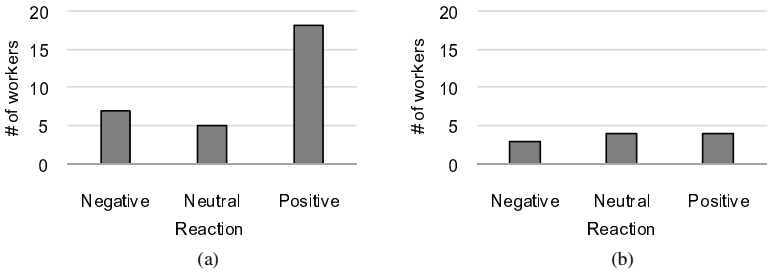


Fig. 7. Effects of the gamification display on the number of contributions for participants who (a) looked at the game metrics vs. (b) never looked at them

staff members and participants. All of the questions were answered within 24 hours. As regards the subjects of the questions, only 1 question asked for linguistic knowledge, while 33 asked for system usage information and 26 asked about proofreading rules.

(R3). H3 could be promising but our results were inconclusive. At least for the limited sample, seniors tended to ask more questions than younger participants.

3.4 Gamification

We added the gamification features to our system in the middle of the target period. Thus we can compare the results without and with the gamification. For analyses, we divided the participants into two groups: those who have looked at the gamification display and those who have never looked at it. We scored the game display as seen when a participant scrolled down the page and the mouse cursor moved over the game display. Fig. 7 shows the number of participants for whom the gamification features had positive, neutral, and negative effects, based on comparing the amount of contributions during each week before and after the introduction of gamification. For the group that looked at the game metrics, the numbers of participants with positive, neutral, and negative reactions were 18, 5, and 7, respectively. For the latter group that never looked at the game metrics, the numbers were 4, 4, and 3. This result seems to show some positive effects of the gamification display. We also compared the effects for senior and young participants and found no apparent effects related to age.

For further analyses, we conducted an online survey. We asked how each component of the gamification display motivated participants with a 5-point Likert scale from 1 (not motivated at all) to 5 (highly motivated). A total of 29 participants (22 senior and 7 young) responded to the survey. The results are summarized in Fig. 8. For the Over-60 group, the average values were 3.3, 3.0, 3.1, 3.6, and 3.6 for ranking, badges, the accumulated number of tasks, the percentage of contribution to the last book, and the number of contributed books, respectively. For the Under-59 group, the values were 3.3, 2.6, 2.6, 3.0, and 3.6. The majority of the respondents most preferred the display of the number of contributed books. In contrast, ranking, badges, and the accumulated number of tasks were least motivational according to self-reporting.

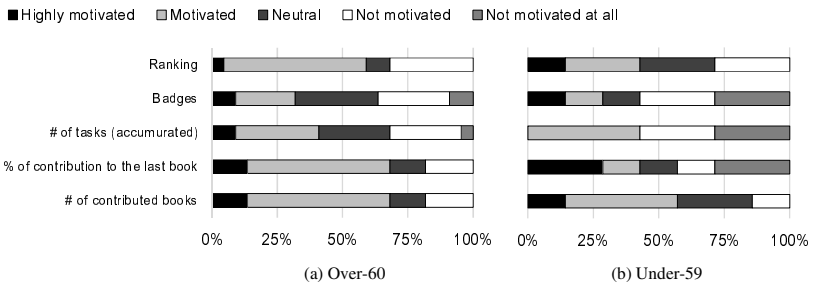


Fig. 8. Subjective evaluation of each gamification component

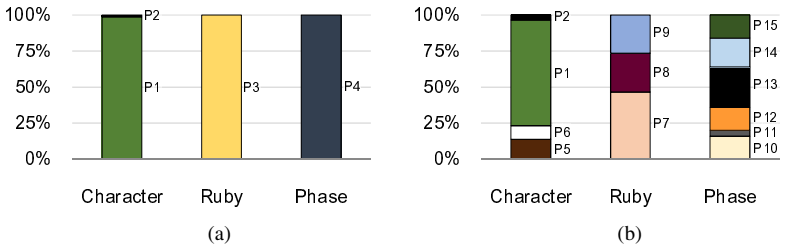


Fig. 9. (a) High-performers for a book vs. (b) long-tail workers accumulating work on a book. Each stacked rectangle represents the amount of the contribution of a worker to the book.

(R4). H4 could be promising but our results were inconclusive. At least for the limited sample, the participants who had looked at the gamification display tended to perform more tasks but no age-related effect was observed. At the same time, the subjective evaluation did indicate that the Over-60 group felt more encouraged than the Under-59.

4 Discussion

4.1 Effects of High-Performance Workers

As shown in Section 3.1, the senior participants tended to complete more work. However, we need to note that the total work performed by the workers was not normally distributed. A few high-performance workers contribute much more than others, and thus apparent tendencies in the total performance are dominated by their outcomes. For some books, almost all of the work was completed by one high-performance worker (Fig. 9-a). However, there are also some books that involved more than 10 workers completing a single book (Fig. 9-b). This observation indicates that the sustainability of micro-tasking community can involve both the contributions of high-performers and long-tail workers. More investigation of the differences among individuals will be needed. Note that the task performance depends not only on the

Table 2. Top X worker's performance by task type

TOP	Character			Ruby			Phrase		
	# tasks	Cumulative total	Age	# tasks	Cumulative total	Age	# tasks	Cumulative total	Age
1	85,269	30.5%	64	9,528	12.2%	49	24,948	18.2%	49
2	37,392	43.9%	57	6,305	20.3%	78	15,198	29.3%	76
3	35,407	56.6%	49	5,270	27.1%	76	10,769	37.1%	45
4	28,956	66.9%	45	5,227	33.8%	64	8,863	43.6%	64
5	27,919	76.9%	61	4,965	40.2%	42	7,525	49.1%	64
10%	6,836	87.6%	–	4,853	46.4%	–	4,426	63.1%	–
25%	957	96.3%	–	1,312	83.3%	–	1,236	86.4%	–
50%	133	99.4%	–	302	95.4%	–	292	97.0%	–

skills of each worker but also on the content and OCR quality of each book. It will be necessary to eliminate the effects of differences among books to assess the skills of individual workers. It is also notable that there were many seniors among the high-performers (Table 2). The involvement of active seniors will be a key to enhancing crowdsourcing applications.

4.2 Complementarity in Competencies

The results in Section 3.2 showed that the differences between senior and younger workers for each task type are observed in terms of trying to start work or not start while the differences are not clear in the performance metrics of the active workers. This may indicate that the dominating factors in the workers' decisions are changed before and after starting work. We had classified the proofreading sub-tasks based on the three competencies of linguistic knowledge, ICT skills, and visual attention. What are the unpredicted competencies? In the experiment, we noticed a fourth competency we called "task-specific knowledge", which may be the most important competency we recognized after starting work. In our context, this is knowledge about the editorial rules in proofreading. In fact, almost half of the questions posted in the Q&A forum asked about editorial rules. It is not hard to anticipate that few of the workers would have such knowledge in the beginning. The lack of task-specific knowledge about editing might mask significant efficiency differences between senior and younger workers, as discussed in Section 3.2.

4.3 Knowledge Transfer among Participants

The Q&A forum was essential for some of the participants, especially among the seniors. It provided them with crucial on-demand support while doing the work. The resulting answers in the Q&A threads also helped other participants with similar questions. In addition, we used the Q&A forum to iteratively update the downloadable instruction manuals. The communications in the forum allowed us to see what information the participants needed and helped us improve our system. However, the

limited results during the experimental period were insufficient to evaluate the age-difference effect on need for Q&A, because there were too few questioners. Another problem for the future is to address the lack of mutual support in the Q&A system. During the experimental period, most of the questions were answered by administrators. One reason might be that most of the questions were related to the system usage or proofreading rules, both of which required expert knowledge that the participants had not yet acquired it. However, since the amount of work that can be done by administrators is limited, mutual support among participants is needed to improve the scalability and sustainability of the micro-tasking system. We are continuing our pilot study and we will examine whether the participants with longer experience are motivated to transfer their knowledge to novices.

4.4 Motivation Encouragement by Gamification

The results indicated that the gamification mechanism has certain positive effects in motivating participants. In addition, the questionnaire results showed that the seniors were more positive about the gamification than the younger participants. More specifically, seniors preferred the gamification components in a general way, while the younger participants favored specific metrics such as the number of contributed books and ranking. Among the gamification components, the visualizations of contributions were most preferred. In particular, the seniors did not like the visualizations of their own efforts or competition with other participants, whereas the younger participants tended to like the competitive aspects. This might be because our experiment involved volunteers who had intrinsic motivations to contribute to society by helping people with disabilities. Other types of tasks may call for other types of incentives. It is future work to examine what types of feedback are effective to encourage senior citizens to participate in other types of micro-tasks such as paid work not for social contribution as well as to assess age differences in the effects of gamification with a larger sample.

5 Conclusion

This paper described some of the characteristics of senior workers observed in a crowdsourcing system for proofreading tasks. The results showed that seniors tended to do more work than young workers. It was indicated that the Q&A and gamification mechanisms are particularly effective for senior workers. Also the potential of the multi-generational approach was shown. The involvement of both senior and younger workers allows gathering different competencies while micro-tasking allows decomposing a larger task into sub-tasks for each competency. Our future work will include individual skill management for performance improvements and individual motivation encouragement for long-term engagement. For skill management, preliminary findings were presented in [12]. For motivation encouragement, basic findings were discussed in this paper. Based on these insights, we will continue investigating elderly participation in crowdsourcing, including different types of tasks, such as paid work.

Acknowledgments. This research was partially supported by the Japan Science and Technology Agency (JST) under the Strategic Promotion of Innovative Research and Development Program. We thank Mr. Sawamura and other staff members of the Japan Braille Library for their support. We also thank all of the participants in our experiment.

References

1. Miura, T., Nakayama, M., Hiyama, A., Yatomi, N., Hirose, M.: Time-Mosaic Formation of Senior Workforces for Complex Irregular Work in Cooperative Farms. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2013, Part II. LNCS, vol. 8010, pp. 162–170. Springer, Heidelberg (2013)
2. DAISY Consortium, http://www.daisy.org/
3. Kobayashi, M., Ishihara, T., Itoko, T., Takagi, H., Asakawa, C.: Age-based Task Specialization for Crowdsourced Proofreading. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2013, Part II. LNCS, vol. 8010, pp. 104–112. Springer, Heidelberg (2013)
4. Bookshare, http://www.bookshare.org/
5. Neudecker, C., Tzadok, A.: User Collaboration for Improving Access to Historical Texts. Liber Quarterly 20(1), 119–128 (2010)
6. Chrons, O., Sundell, S.: Digitalkoot: Making Old Archives Accessible Using Crowdsourcing. In: Proc. HCOMP 2011 (2011)
7. Marmot, M., Banks, J., Blundell, R., Lessof, C., Nazroo, J. (eds.): Health, wealth and lifestyles of the older population in England: ELSA 2002. Institute for Fiscal Studies (2003)
8. Ouchi, Y., Akiyama, H. (eds.): Gerontology – Overview and Perspectives, 3rd edn. Univ. of Tokyo Press (2010) (in Japanese)
9. Staffing Mature Worker Survey, http://www.goldenworkers.org/images/publication/mature_workers_survey_2012_adecco.pdf
10. Leung, R., Tang, C., Haddad, S., McGrenere, J., Graf, P., Ingriany, V.: How Older Adults Learn to Use Mobile Devices: Survey and Field Investigations. ACM Trans. Access. Comput. 4(3), Article 11 (2012)
11. Czaja, S.J., Lee, C.C., Branham, J., Remis, P.: OASIS Connections: Results From an Evaluation Study. The Gerontologist 52(5), 712–721 (2012)
12. Kobayashi, M., Ishihara, T., Kosugi, A., Takagi, H., Asakawa, C.: Question-Answer Cards for an Inclusive Micro-Tasking Framework for the Elderly. In: Kotzé, P., Marsden, G., Lindgaard, G., Wesson, J., Winckler, M. (eds.) INTERACT 2013, Part III. LNCS, vol. 8119, pp. 590–607. Springer, Heidelberg (2013)
13. Farzan, R., DiMicco, J.M., Millen, D.R., Dugan, C., Geyer, W., Brownholtz, E.A.: Results from deploying a participation incentive mechanism within the enterprise. In: Proc. CHI 2008, pp. 563–572. ACM (2008)
14. McCallum, S.: Gamification and serious games for personalized health. Stud. Health Technol. Inform. 177, 85–96 (2012)
15. Takagi, H., Harada, S., Sato, D., Asakawa, C.: Lessons Learned from Crowd Accessibility Services. In: Winckler, M. (ed.) INTERACT 2013, Part I. LNCS, vol. 8117, pp. 587–604. Springer, Heidelberg (2013)
16. Web Content Accessibility Guidelines (WCAG) 2.0, http://www.w3.org/TR/WCAG20/