Testing the Wisdom of Crowds in Students' Group Work

AVINASH MALLA
Oulu University of Applied Sciences,
School of Business and Information Management, Oulu, Finland

TEPPO RÄISÄNEN
Oulu University of Applied Sciences,
School of Business and Information Management, Oulu, Finland

Abstract: In his book, "The Wisdom of Crowds", James Surowiecki argues that a group becomes "smart" and makes better decision than an individual if the group satisfies the four conditions-diversity of opinion, independence, decentralization and aggregation. This paper intends to investigate whether groups that have higher diversity of opinion, independence, decentralization and aggregation perform better than the groups that don't. Ideally, the groups should perform better when they satisfy the four conditions better. One set of groups followed The Wisdom of Crowds hypothesis while another set of groups, working on a different task, did not. Both sets of groups, however, performed better when they had higher diversity of opinion.

Keywords: wisdom of crowds, knowledge management, diversity of opinion, independence, decentralization, aggregation.

1 Introduction

While visiting a country fair in Plymouth, England, a scientist named Francis Galton came across a weight judging competition where participants were asked to guess the weight of an ox. All the guesses were then combined to calculate the average guess of the participating crowd. He discovered that the crowd's guess (1197 lbs) was surprisingly close to the actual weight (1198 lbs). This estimate of the crowd was closer to the actual weight than any individual guesses [1]. In his book, The Wisdom of Crowds, James Surowiecki examines this behaviour of groups. He studies the collective wisdom of crowds and the conditions necessary to utilize this collective wisdom. Surowiecki argues that under the right circumstances, groups are remarkably intelligent and often smarter than the smartest people in them. According to him, there are four conditions that characterise the wise crowd: diversity of opinion, independence, decentralization and aggregation [1]. This study intended to test whether a group fulfilling these conditions would actually perform better than a group that does not.

This study was conducted during the "Innoweek" event at the School of Business and Information Technology at the Oulu University of Applied Sciences. For this annual event, students are divided into groups and these groups are challenged to come up with a solution to a real-life problem provided by external commissioners. The questionnaire was distributed on the last day of this week-long event and asked them whether their group fulfilled the four conditions as stated by Surowiecki.

The research was used to rank the groups based on the extent to which each group satisfied the four conditions of Wisdom of Crowds. In other words, the research ranked the teams based on their smartness. This ranking was compared with the final results of Innoweek to check whether the smartest groups actually performed better at Innoweek.

2 Wisdom of Crowds

Aristotle, a 4th century BC Greek philosopher, is known as the first person to discuss collective wisdom in his book Politics.

> *For the many, of whom each individual is but an ordinary person, when they meet together may very likely be better than the few good, if regarded not individually but collectively, just as a feast to which many contribute is better than a dinner provided out of a single purse. [2]*

Apart from the Francis Galton experiment, Surowiecki uses another example of the Google PageRank algorithm to further explain how Google uses the concept of collective wisdom to return the best search result at lightning speed. The algorithm largely depends on aggregating the links to one website from another, where each link is considered as a vote. Using these and more case examples, Surowiecki argues that a group becomes "smart" and makes a better decision than an individual, provided that the group satisfies the four conditions - diversity of opinion, independence, decentralization and aggregation [1]. He uses the example of prediction markets to exhibit the application of Wisdom of Crowds. In the prediction markets designed by the Iowa Electronic Markets (IEM), to predict the election results, people could trade futures contracts based on how they think a candidate will perform in an election. The IEM predictions were found to be accurate three-fourths of the time [1].

A group will not perform any different than an individual if group members have similar opinions. Surowiecki argues that each person should have some "private" information that is unique. Diversity of Opinion can be maintained by assembling a diverse group of people who possess varying degrees of knowledge and insight. The group should contain experts as well as people without knowledge in the field [1]. In large groups, diversity is obvious, but in smaller groups, it should be encouraged to avoid influence [1]. Diversity here should be understood as cognitive diversity, where people have various degrees of knowledge and insight [1].

Cohesion in homogeneous groups is more than in diverse groups. The group members' dependence on the group increases as they become more cohesive. This consequently insulates them from external influence and convinces them that the group's judgment must be right [1]. Diversity not only adds different perspectives to the group but also facilitates people to speak their minds [1]. Surowiecki gives an example of how the Kennedy administration failed during the Bay of Pigs Invasion due to a lack of opposing opinions as it was carried out under the guidance of the administration's few like-minded people [1].

Independence among group members is another condition for the group to be smart. The smartest groups consist of people with diverse opinions who are not influenced by others. Necessary measures should be taken to avoid groupthink. Groupthink occurs when a group makes faulty decisions because group pressures lead to deterioration of "mental efficiency,

reality testing, and moral judgment" [3]. The people in the group should be able to make their own minds and answer the questions independently and uninfluenced by groupthink.

According to Surowiecki, independence is important due to two reasons. First, it ensures that the human mistakes don't correlate. Second, independent individuals bring new data rather than the familiar old information [1].

Decentralization is the third condition for a group to be smart. Decentralization means that the group members belong to different backgrounds and are thus specialized in different areas and can draw on their local knowledge when making decisions. This contributes to diversity and independence [4]. According to Surowiecki, a group can become smarter by including people with local and specific knowledge. The chances of having a good solution increases when a person is closer to the context. Decentralization facilitates specialization, which in turn feeds decentralization. Specialization increases the productivity and efficiency along with widening the scope and fostering diversity of opinion and information gathering. Decentralization is of paramount importance to tacit knowledge. Tacit knowledge is knowledge that can't be easily summarized or conveyed to others because it is specific to a particular place or job or experience, but it is nonetheless tremendously valuable [5]. This definition sheds light on the assumption that the closer a person is to a problem, the more likely he or she is to have a good solution to it [1].

Aggregation is the method of converting various private judgements of individuals in a group into a collective output [1]. It explores whether people are making individual decisions or a collective one. Surowiecki identifies aggregation as the fourth condition for a group to be smart. He gives the example of the Linux Operating System, where a large number of coders working individually in a decentralized fashion contribute to improvement of the system. These contributions are then scrutinised by a small group of coders including the creator Linus Torvalds before being implemented, thus aggregating the work done by numerous autonomous individuals [1].

According to Lyon & Pacuit, mathematical aggregation methods like mean, median, mode, etc. can be used to aggregate the crowd's contribution. Group deliberation is another method of aggregation, which entails the crowd meeting to discuss the problem at hand before arriving at a collective judgement [6]. Prediction markets is another commonly used aggregation method to aggregate opinions, and it has been more popular lately [1]. Wikipedia articles and Facebook homepage can be viewed as examples of aggregation [7]. Facebook's algorithm gathers updates from the user's friends, aggregates them and only shows updates it deems important to the user.

3 Case Innoweek

Innoweek is an annual event organized by the School of Business and Information Technology at Oulu University of Applied Sciences (referred from now on as OUAS). Students from all the degree programs of the school participate in the event. They are awarded three credits for this project work course. The supervising teachers divide the students into groups. The teachers give special attention to maintaining the diversity by ensuring students from various degree programs and of various nationalities are chosen while forming groups. This idea of group work and dynamics of diversity in the group provide an opportune premise to investigate the proposition of the wisdom of crowd.

3.1 Innoweek

The objective of Innoweek as stated in the OUAS curriculum "*enables the students to act in the innovation process to develop activities of organizations or to identify new business opportunities. The goal is to develop innovation and team working skills.*" [8]. The students work in groups to tackle real life business problems. At the end of the week-long event, the groups present their ideas in front of a company representatives. In this year's event, two organizations (commissioners) challenged the students with their problems. The first commissioner was PrintoCent – a company that creates innovative products and business solutions based on printed electronics. They challenged the students to come up with proposals and marketing plans for their various products. The other commissioner, Student Union of Oulu University of Applied Sciences (referred from now as OSAKO), challenged the students to propose a plan to increase the voter turnout in the student union elections. The students had one week to work in their group to prepare a presentation that would be presented on the final day of the event.

Before the week of the event, interested students enrolled in the course. Students from all the degree programs at OUAS were eligible and a total of 61 students enrolled in the course. The supervising teachers for the course then assigned the students to two groups - PrintoCent (Group A/Task 1) and OSAKO (Group B/Task 2). The students from each task were then divided into smaller groups of four to five students. Eight groups participated in the first task while the second task had six. To maintain diversity while forming groups, the teachers chose students from different degree programs and included at least one foreign student in every group. On the first day of the event, the commissioners presented the challenge to their corresponding student groups. The students worked in their groups each day of the week to come up with the best solution. On the final day of the week, the groups presented their ideas to the commissioners.

3.2 Research Methodology

The objective of this research was to find out if the groups participating in Innoweek adhered to the theory of the wisdom of crowds as defined by James Surowiecki. The research intended to answer the following research questions:

1. Do the groups follow the wisdom of crowds?
2. What is the correlation between the four conditions - diversity of opinion, independence, decentralization and aggregation - and their performance? Do the groups that satisfy these four conditions better actually perform better in the tasks?
3. Is any of the four conditions more important than the others?

To answer these questions, a questionnaire was prepared and distributed to the students on the final day of Innoweek. The questionnaire was divided into four sections based on the four conditions of wisdom of crowds and each section contained three statements. The participants were asked to rate their opinions about the statements using a five-level Likert scale. Each level carried a score from 1 to 5. Every participant's average score for each of the four conditions was calculated. Each group's average score was then calculated by aggregating the average scores of the participants to derive the groups' average scores for Diversity of Opinion, Independence, Decentralization and Aggregation. The groups' average score for each condition was again aggregated to arrive at the final Wisdom of Crowds (WoC) score.

4 Results

This section analyses the outcome of the research by comparing each condition of Wisdom of Crowds with the actual performance of the groups. The correlations between each condition and the actual performance of groups in both tasks are compared. Out of 61 students who enrolled in Innoweek, 46 participated in the survey. Even though the course was open for students from all the degree programs, the business students outnumbered the others with 20 from the Degree Program in International Business and 17 from *Liiketalouden ammattikorkeakolututkinto* - the Bachelor of Business Administration program taught in Finnish.

4.1 Definitions

During the research, various terminology was created to explain the calculations. These terms and their mathematical meanings are explained in this section.

Actual Rank: The actual rank was derived from the results of Innoweek. The participating groups were ranked from first to last based on the scores they received from the judges.

Diversity of Opinion Rank: Ranking of the groups from first to last based on the Diversity of Opinion score.

Independence Rank: Ranking of the groups from first to last based on the Diversity of Opinion score.

Decentralization Rank: Ranking of the groups from first to last based on the Decentralization score.

Aggregation Rank: Ranking of the groups from first to last based on the Aggregation score.

Wisdom of Crowds (WoC) Rank: Ranking of the groups from first to last based on the overall WoC score. WoC score was calculated by averaging the scores of the four conditions.

Task 1: The task commissioned by Commissioner 1

Task 2: The task commissioned by Commissioner 2

4.2 Correlations

In order to establish the relation, if any, between the four conditions of wisdom of crowds and the actual performance of the groups, Spearman's Rank Correlation was used. Spearman's Correlation is used to measure the degree of correlation between two sets of observations or between paired values when the relative order of magnitude is given for each series [9]. The scores for each of the four conditions were ranked from first to last and compared with the actual rank using the Spearman's Rank Correlation Coefficient formula.

$$\rho = 1 - \frac{6\Sigma\, d^2}{n(n^2-1)} \tag{1}$$

where, n = number of items being ranked, d = the numerical difference between corresponding pair of ranks, (Xi - Yi) [10].

The value for Spearman's Coefficient (also known as Spearman's rho and denoted by 'ρ') ranges between -1 to +1. When ρ is >0, it indicates the positive correlation; when ρ is <0, it indicates the negative correlation in the order of ranking or selection or judging. The strength

of the Spearman's Correlation Coefficient can be interpreted as follows using the absolute value of ρ:

4.2.1 Diversity of Opinion versus Actual Rank

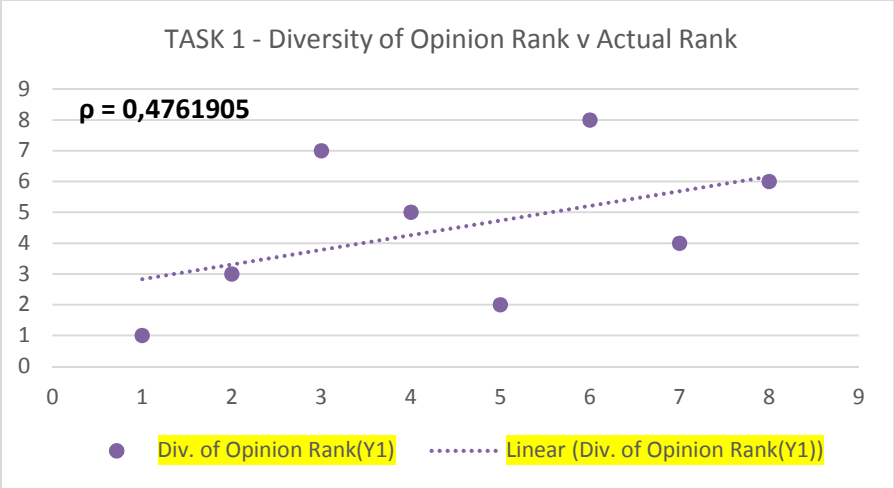

Fig. 1 *Diversity of Opinion versus Actual Rank for Task 1*

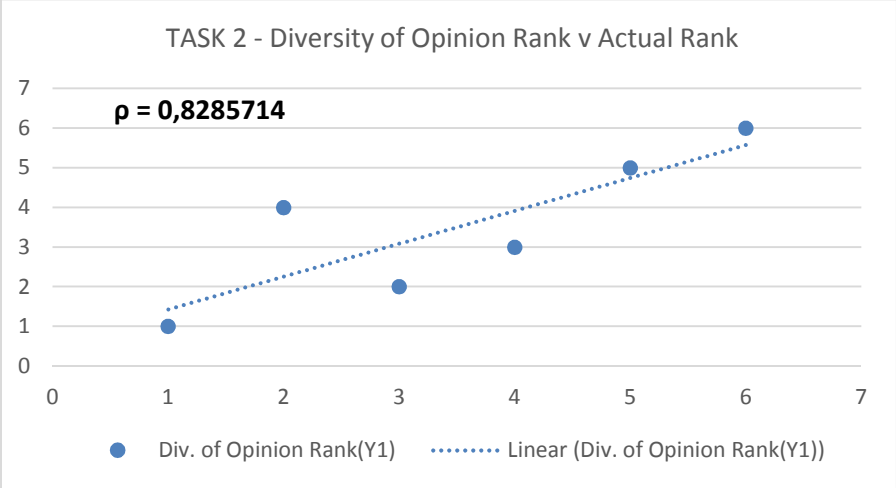

Fig. 2 *Diversity of Opinion versus Actual Rank for Task 2*

These graphs exhibit the correlation between the Diversity of Opinion and the Actual Rank in the two tasks. The positive value of ρ indicates the positive correlation between the Diversity of Opinion and the actual performance of the groups. The groups in both tasks exhibited increase in the actual performance when there was increase in the diversity of opinion. Task 1 with the ρ value of 0,4761905 (see Fig 1) indicates a moderate positive correlation while Task 2 displays a very strong positive correlation with ρ value of 0,8285714 (see Figure 2). This suggests that the groups with higher diversity of opinion performed better in their task. The teachers supervising the students during Innoweek tried their best to make sure the students came from different degree programs and nationalities while forming groups. This heterogeneity of backgrounds ensured the diversity of opinion in groups. Thus, the more diverse groups evidently performed well in their task.

4.2.2 Independence versus Actual Rank

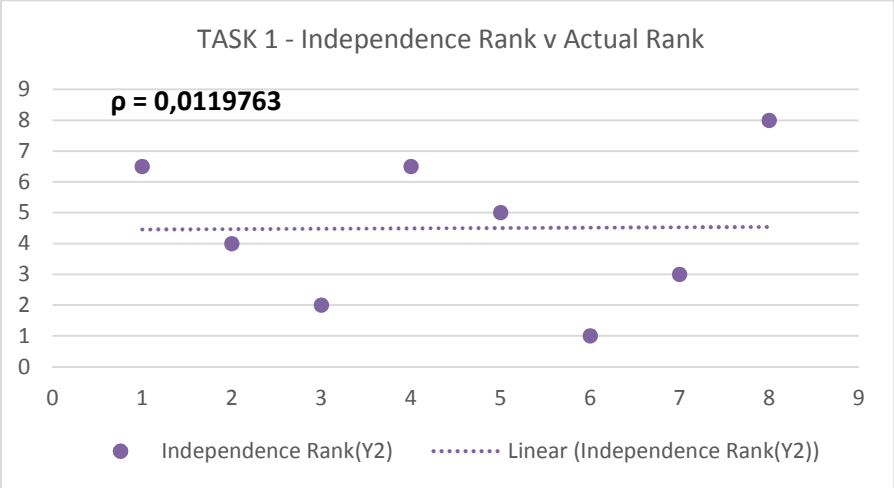

Fig. 3 *Independence versus Actual Rank for Task 1*

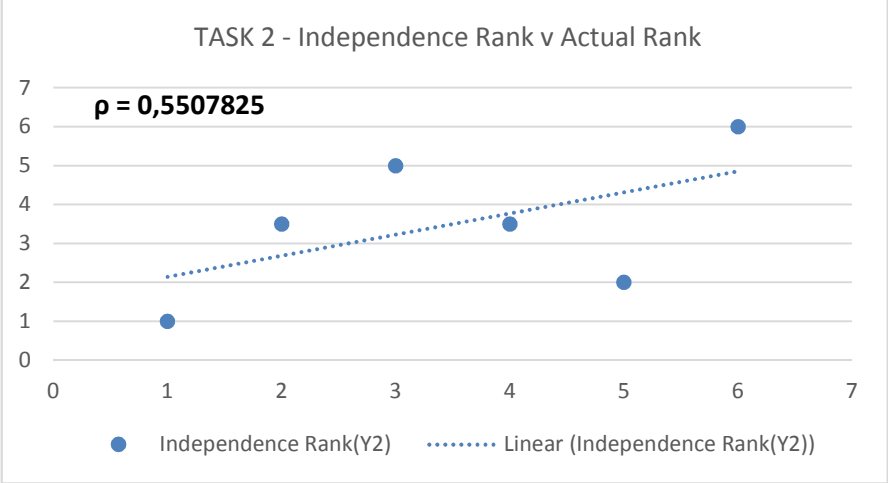

Fig. 4 *Independence versus Actual Rank for Task 2*

These graphs demonstrate the correlation between independence and performance of the task. The Spearman's Correlation Coefficient for Independence Rank versus Actual Rank for Task 1 is 0,0119763 (see Figure 3). This suggests that there is a very weak positive correlation between the two. Ideally, the more independent the group members are, the better the outcome of the task. It can be observed that the groups in Task 2 adhere to the postulation of Wisdom of Crowd more than the groups in Task 1.

4.2.3 Decentralization versus Actual Rank

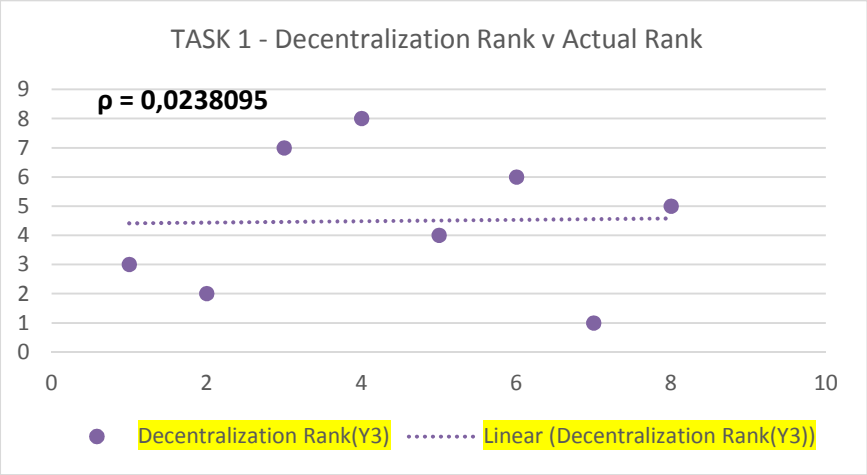

Fig. 5 *Decentralization versus Actual Rank for Task 1*

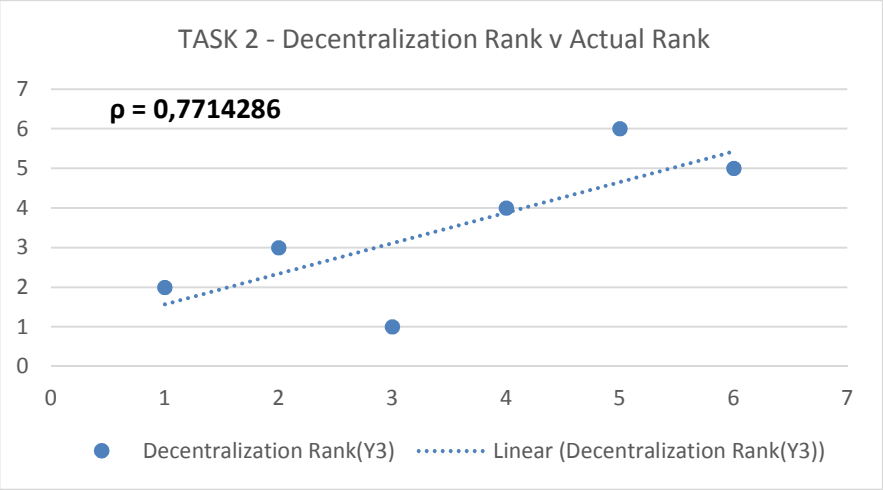

Fig. 6 *Decentralization versus Actual Rank for Task 2*

These graphs demonstrate the correlation between independence and performance of the task. The Spearman's Correlation Coefficient for Independence Rank versus Actual Rank for Task 1 is 0,0119763 (see Figure 3). This suggests that there is a very weak positive correlation between the two. Ideally, the more independent the group members are, the better the outcome of the task. It can be observed that the groups in Task 2 adhere to the postulation of Wisdom of Crowd more than the groups in Task 1.

4.2.4 Aggregation versus Actual Rank

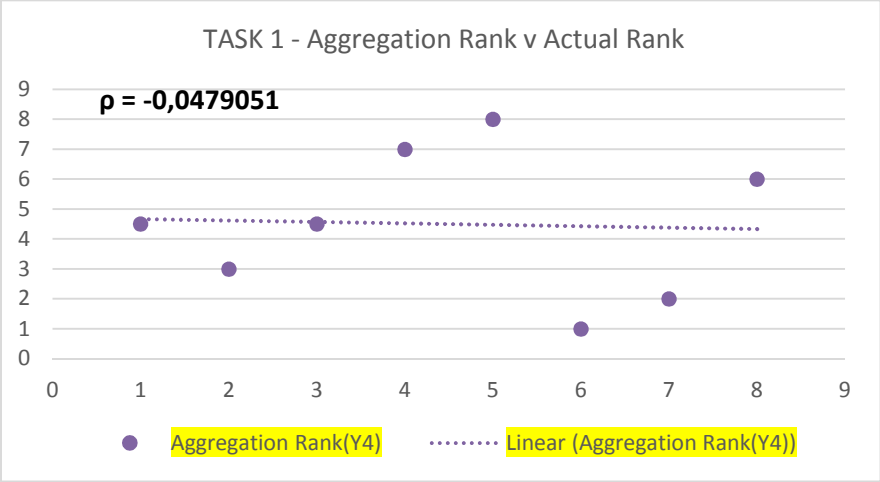

Fig. 7 *Aggregation versus Actual Rank for Task 1*

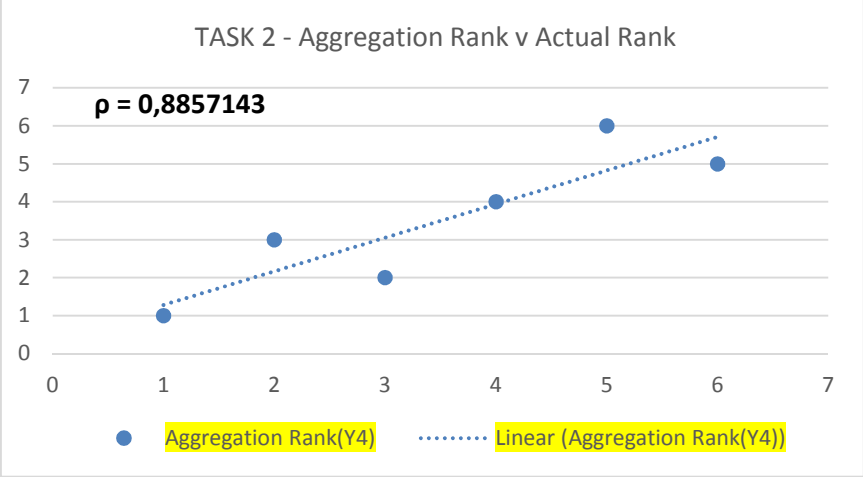

Fig. 8 *Aggregation versus Actual Rank for Task 2*

When the Aggregation Rank is compared with Actual Rank using the Spearman's Rank Correlation Coefficient, the groups in Task 1 display a very weak negative correlation with the ρ value of -0,0479051 while the groups in Task 2 demonstrate a very strong positive correlation with the ρ value of 0,8857143 indicating better performance by the group that exercised higher level of participative decision making.

5 Conclusion & Discussion

James Surowiecki argues that when a group of individuals working together satisfy the four conditions: diversity of opinion, independence, decentralization and aggregation, they perform better than an individual. Through his book "The Wisdom of Crowds", upon which this study is based, Surowiecki uses real life examples to show that the average opinion of groups is often more correct than most individuals in the group. This hypothesis was examined on the students working in groups for Innoweek. The results showed that the groups working on one task abided by the hypothesis while results of the groups working on the other task did not comply.

This study was built upon the premise that a group performs better if the four conditions of diversity of opinion, independence, decentralization and aggregation are met. When we examined the hypothesis on the groups in Innoweek, two contrasting results appeared. The groups performing one task did better only when they had higher diversity of opinion while the groups performing another task did better when they had higher diversity of opinion as well as higher independence, decentralization and aggregation. This means that the groups performing the second task followed the Wisdom of Crowds but the ones performing the first task did not.

The conflicting outcome of the research may be due to certain factors pertaining to the nature of the tasks and statistical shortcomings. The first task challenged the students to come up with plans to commercialize a company's various products. This might have affected the independence among group members as their scope of performance was in some way limited to the products' features and the company's guidelines. The second task, on the other hand, challenged students to come up with a plan to increase the voter turnout during OSAKO's annual student union elections. This was more of an open task, where the participants' scope was not limited, enabling them to come up with any kind of ideas.

The four conditions were examined individually against the actual performance using the Spearman's Rank Correlation Coefficient, where each group's scores were ranked from first to last and compared with the actual results of Innoweek. This rank versus rank correlation displayed the effect any or all of the four conditions had on the actual performance. According to the results, the groups working on the task commissioned by PrintoCent only performed better when there was higher diversity of opinion. These groups did not perform better when they had higher independence, decentralization or aggregation. On the other hand, the groups working on the OSAKO's commissioned task performed better when they had higher diversity of opinion, independence, decentralization or aggregation. Since Diversity of Opinion is a common theme of better performance of the groups, it could be said that groups perform better when members of the groups contribute with unique information and perspective they have.

The research was limited by various factors. The low number of participants affected the quality of data hindering the chances to attain optimum heterogeneity in groups. For future research, the experiment can be conducted in an organizational setting with a much larger group of people, thus ensuring higher diversity of opinion, independence and decentralization. A more scientific approach should be taken when measuring the conditions of wisdom of crowds and assessing the performances. The same experiment can be repeated multiple times with different sets of data to obtain any conclusive outcome.

Innoweek, being a school task, was in some manner limited to various parameters enforced by the course and thus hindering the independence factor among the participants. All the participants of the survey were students. The grading scale of the course was pass or fail. Working harder than others did not ensure a better outcome for them. This may have

discouraged some students from performing at their optimum level. The methodology used to collect data for the study might also have been unscientific. The survey questions might not have been able to provide a true view of the groups' performance as it was completely dependent on the participants' opinions. Response bias could have also affected the results. Since they were asked to score their own work during the survey, they might have rated themselves higher.

The two tasks were judged by their corresponding commissioners. The judges might have had differing judgement criteria. The lack of a scientific mechanism for the judges to evaluate the participants' performance and consequent subjective judgement by them could also have affected the outcome. Another reason could be the number of participants and groups. The first task had more participants than the second task. About 85% of the participants working on the first task participated in the survey while only 69% of those working on the second task answered the survey. This disparity in the number of participants might also have affected the outcome of the study. Inclusion of qualitative data could have improved the results. Interviewing some participants could have given a detailed perspective on the nature of the tasks allowing better understanding of the disparity in the results.

References

1. SUROWIECKI, J., 2005. *The Wisdom of Crowds*. 1st ed. New York: Anchor Books.
2. ARISTOTLE, 350 B.C.E. *Politics*. Translated by Benjamin Jowett., 1999. Kitchener: Batoche Books. (Irvine, 1972).
3. JANIS, IRVING L., 1972. *Victims of Groupthink*. 1st ed. New York: Houghton Mifflin.
4. RÄISÄNEN, T., 2013. *Investigating aggregation and diversity for wisdom of crowds approach*. 8th International Workshop on Knowledge Management, October 18-19, Bratislava, Slovakia.
5. HAYEK, F. A., 1945. *The Use of Knowledge in Society*. American Economic Review 35. (Hayek, 1945).
6. LYON, A., & PACUIT, E., 2014. *The Wisdom of Crowds: Methods of Human Judgement Aggregation*. Cited 15.07.2016, http://pacuit.org/media/mypapers/woc-handbook.pdf
7. RÄISÄNEN, T., & WATTS, S., 2012. *Experimenting with Wisdom of Crowds as a Source of Knowledge.* The 9th International Conference on IT Applications and Management, December 20-21, Yonsei University, Wonju, Korea. 4.
8. Oulu University of Applied Sciences, *Innoweek Innovation Project*. Internal source. https://oiva.oamk.fi/tietoa_opiskelusta/opintojen_suunnittelu/opas/koulutusohjelmat/?koulutus=bit2016s&lk=s2016&alasivu=opintojakso&kieli=en&oj=K1115YY_en [Accessed 18 October 2016].
9. CHIKKODI, C.M. & SATYAPRASAD, B.G. 2009. Business Statistics. Mumbai, India: Himalaya Publishing House. ProQuest ebrary [Accessed 23 November 2016].
10. KVAM, P. H., & VIDAKOVIC, B., 2007. *Nonparametric Statistics with Applications to Science and Engineering*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Contact data:

Avinash Malla
Oulu University of Applied Sciences
P.O. Box 222
FI-90101 OULU, Finland
malla.avi@gmail.com

Teppo Räisänen, Principal Lecturer
Oulu University of Applied Sciences
P.O. Box 222
FI-90101 OULU, Finland
Teppo.Raisanen@oamk.fi