A Universal Intelligence Measurement Method Based on Meta-analysis

Zheming Yang[*°] and Wen Ji[*†]
[*] Institute of Computing Technology, Chinese Academy of Sciences
[†] Peng Cheng Laboratory
[°] University of Chinese Academy of Sciences
E-mail: jiwen@ict.ac.cn

Abstract— **The multiple factors of intelligence measurement are critical in the intelligent science. The intelligence measurement is typically built at a model based on the multiple factors. The different digital self is generally difficult to measure due to the uncertainty among multiple factors. Effective methods for the universal intelligence measurement are therefore important to different digital-selves. In this paper, we propose a universal intelligence measurement method based on meta-analysis. Firstly, we get study data through keywords in database and delete the low-quality data. Secondly, after encoding the data, we compute the effect value by Odds ratio, Relatve risk and Risk difference. Then we test the homogeneity by Q-test and analysis the bias by funnel plots. Thirdly, we select the Fixed Effect and Random Effect as statistical model. Finally, simulation results confirm that our method can effectively solve the multiple factors of different digital self. Especially for the intelligence of human, machine, company, government and institution.**

I. INTRODUCTION

A. Background

With the popularization of intelligent science, the application of intelligence measurement become more and more extensive. For example, Gignac and Gilles E [12] found the intelligence measurement can moderate the effect between brain volume and intelligence. Vamsi V [13] measure the business intelligence (BI) by adopting the IT-based performance measurement systems (PMS) to evaluate the performance of organization. Cengiz Kahraman [14] measure the collective intelligence to evaluate the performance in energy systems. Therefore, intelligence measurement has brought many changes to our lives.

Now intelligence measurement methods can be divided into human IQ test, machine intelligence measurement and universal intelligence measurement. In IQ test, it measures an individual intelligence mainly through their perception and understanding of knowledge, words and graphics. At present, the two mainstream IQ tests in the world are the Binet-Simon intelligence Scale and the Wechsler Intelligence Scale. They both measure people's intelligence by answering many questions. MC Mcgrath [6] defined standard score as a type of normally distributed standard score (with a mean of 100 and a standard deviation of 15) that represented level of performance on tests of cognitive ability. In the machine intelligence measurement, it is mainly based on Turing test. Turing A M [1] adopted the mode of "question" and "answer" in 1950, that is, the observer talks to two test subjects by controlling the typewriter, one is a person, the other is a machine. He measures the intelligence of the machine by the questions that the observer constantly raises. Cochrane P [2] assumed an entropic measure able to account for the reduction or increase in the system information or state change, before and after the application of intelligence. Then he defined the machine intelligence as the change in entropy. Legg S [15] take a number of well -known informal definitions of human intelligence that have been given by experts, and extract their essential features. These are then mathematically formalized to produce a general measure of intelligence for arbitrary machines. Zeungnam Bien [16] analyzed those engineering systems or products that are said to be intelligent and have extracted four common constructs. Then they adopted the Sugeno fuzzy integral and the Choquet fuzzy integral to find a number called machine intelligence quotient. In the universal intelligence measurement, Hernandez Orallo J proposed C-test [3],[4] in 2000, which can calculate many useful test problems. And these questions have been proved to be related to real IQ test scores. Then in 2010, based on Kolmogorov's complexity, C-test and Compression-enhanced Turing test. Legg S and Insacabrera J proposed a universal idea of intelligence measurement [9] in anywhere and anytime and defined the universal intelligence [10]. Vaibhav Gavane proposed a new measurement [17] of intelligence for general reinforcement learning agents, based on the notion that an agent's environment can change at any step of execution of the agent. And the resulting intelligence measurement is more general than the universal intelligence measurement [9] and the anytime universal intelligence test [10].

However, according to the results of these papers, all the proposed methods have some drawbacks. Although they draw into environmental complexity and time, they don't consider the relationship between the multiple factors. So, they can't combine the different digital self. Therefore, the multiple factors and the heterogeneity are main difficulties in universal intelligence measurement. We propose a universal intelligence measurement method based on meta-analysis to solve the problem. Meta-analysis originates from statistics. It is a statistical method that integrates multiple research data. It can conduct a unified integrated analysis of the existing conclusions, and objectively evaluate the existing research data to draw more valuable conclusions. At present, meta-analysis has been widely used in medical field [7], [8], social science field [21],[20] and library information science field [19],[18]. Myszkowski N analyzed the relationship between intelligence and visual measurement [11] by meta-analysis. On the base of these, we solve the problem of multiple factors. And it combines human, machine, company, government and institution at the same time, as shown in Figure 1.

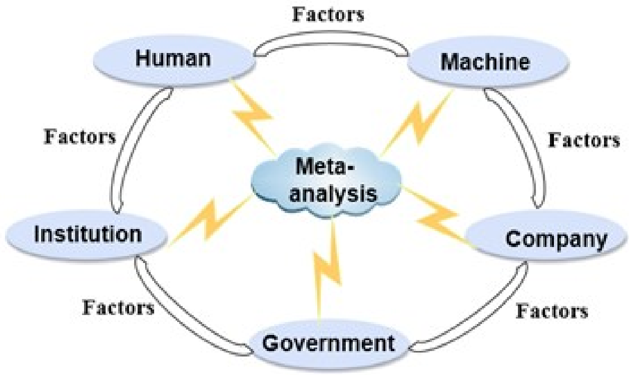

Fig.1 The digital self of meta-analysis

B. Contribution

In this paper, we propose a universal intelligence measurement method based on meta-analysis. The contribution of this paper is following:

1. We consider the relationship between the multiple factors by meta-analysis.

2.Our method solved the heterogeneity by studying many different data. And it can combine the different digital self. Especially for the intelligence of human, machine, company, government and institution.

3.We first apply meta-analysis for intelligent science. It provides a great idea for other scholars.

The rest of the paper is organized as follows. In Section 2, we introduce the meta-analysis method and concrete implementation steps. Then we introduce the method of merger effect value, Q-test and bias analysis. The experimental results are provided in Section 3. Finally, the conclusion and future work are presented in Section 4.

II. UNIVERSAL INTELLIGENCE MEASUREMENT METHOD

A. Construction of Data Set

Retrieving data of Meta-analysis is different from traditional retrieval method. It should retrieval as much research data as possible related to intelligent measurement. It is necessary to provide a large number of key words and a retrieval database for meta-analysis. Then we retrieval the keywords in the database to get data set. By researching the current academic progress of intelligence measurement, we determined the keywords and database as: Key words: Intelligence, Measurement，Universal, Increment, Crowd，Level, Digital, Physical, Crowd,Network,Entropy,Machine,Artificial. Database: Google Scholar. Finally, we get a total of 42 papers that cover all fields related to intelligence measurement [5].

There may be some low-quality data in the data set. Therefore, we established a data filtering standard to delete low-quality data. The data filtering standard is depending on the research subject and the research data. In this paper, we determine the data filtering standard as: If the data title contains anyone of the 'Intelligence' or 'Measurement', we regard it as high-quality data. And if it does not contain the above two keywords, but contains more than two other arbitrary keywords, we also regard it as high-quality data. In addition, the filtering standard is not fixed and can be adjusted according to the actual situation. For example, it can also be regarded as high-quality data as long as the research is highly relevant to the title. Finally, we selected eight papers as the data set of meta-analysis. As shown in Figure 2.

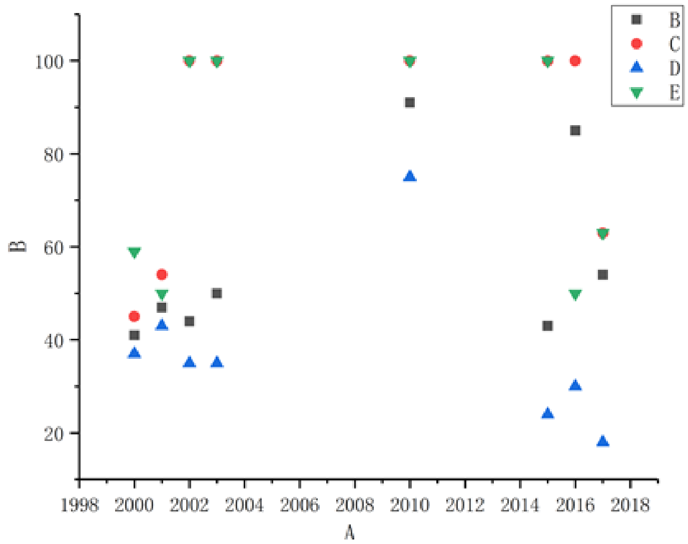

Fig.2 The distribution of data set

We encode the data set for the statistical analysis. The encoding format is as follows: Number-Author-Time. As shown in Table I. Then the coded papers are put into the data set in turn. In addition, the size of the number only represents the order of coding.

Table I. The Coded data.

Number	Author	Time
1	John Duncan	2000
2	Hee-Jun Park	2001
3	Zeungnam Bien	2002
4	Jacob W. Crandall	2003
5	José Hernández-Orallo	2010
6	Hao Zhong	2015
7	Jose Hernandez-Orallo	2016
8	Monireh Dabaghchian	2017

B. *Calculation of Effect Value*

Effect value is one of the most important factors in meta-analysis. Meta-analysis needs to turn multiple results into a unified statistical factor of effect value because they are heterogeneous. In order to solve the problem that the coefficients of factors are different, we select some statistical variable according to the particularity of intelligence measurement. In this paper, we select the OR (Odds ratio), RR (Relatve risk) and RD (Risk difference) as effect values.

C. *Homogeneity Test*

Homogeneity test is to test the rationality of merging results in data sets. It is mainly to check whether the results of every data can be merged or not. In this paper, we use the Q-test to test the homogeneity. The Q-test obeys the chi-square distribution with degree of freedom k-1, where k is the number of effect values. If Q is statistically significant, means these effects values are heterogeneous distributions. We should adopt Random effect model because it can consider the variation between studies and estimate the average of effects distribution at the same time. Then it avoids underestimating the weight of small samples or overestimating the weight of large samples. It can also get a larger confidence interval and then obtain a better conclusion. If Q is not statistically significant, the results of fixed effect model and random effect model are similar. But if the statistical factor of Q-test is near the critical value, two models should be used simultaneously. Finally, we compare the difference of parameter estimation.

The steps of Q test are as follows:

(1) Arrange the data in incremental order: X1, X2, X3, …, Xn-1, Xn.

(2) Calculate the D-value between the maximum and minimum: Xmax-Xmin.

(3) Calculate the absolute value of the D-value between the suspicious value and its nearest data.

(4) Calculate Q (Q equals the D-value in (3) divided by the D-value in (2)).

(5) Values are obtained by looking up tables based on the number of measurements N and the required confidence level (e.g. 95%).

(6) Judgment, if Q > the value of Q in the table, then discard the suspicious value, otherwise it should be retained.

D. *Bias Analysis*

In this paper, we select the method of funnel plotting to analysis the bias. The bias analysis is mainly the accuracy of each effect value increase with the sample size. We take the effect value as abscissa and the standard error as ordinate to plot. If there is no bias, it should be an inverted funnel. And the points on funnel plot are symmetrically dispersed around the real value of the point estimate of the effect value. The standard errors of small samples are large and scattered at the bottom of funnel plot. With the increase of sample size, the accuracy is also increased and the scatter points are more concentrated. On the contrary, there are bias problems.

III. EXPERIMENT

The experimental environment of this paper is completed under the RevMan 5.3. After importing data set, we select the binary variables as data types and select Mantel-Haenszel as analysis methods. And we select OR (Odds ratio), RR (Relatve risk) and RD (Risk difference) as effect values and select Fixed Effect and Random Effects as statistical models. Finally, we analysis their advantages and disadvantages. Figures 3-8 are the results of the experiment. The center of rectangle represents the point estimate of effect value. The length of rectangle represents the confidence intervals of effect value. And the larger the confidence interval of the effect value, the less accurate the result is.

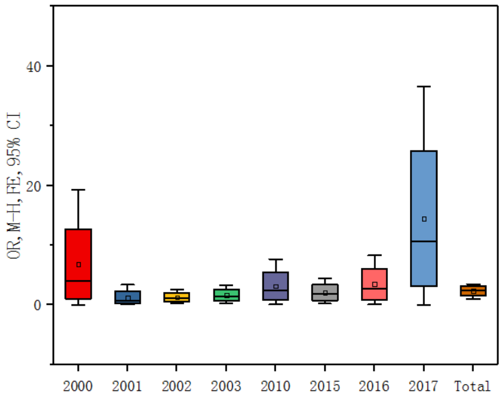

Fig.3 The Box-plot of OR-Fixed

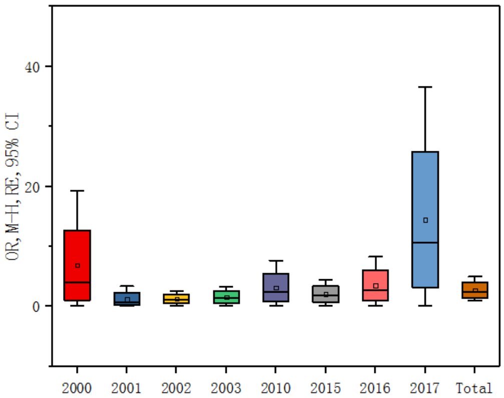

Fig.4 The Box-plot of OR-Random

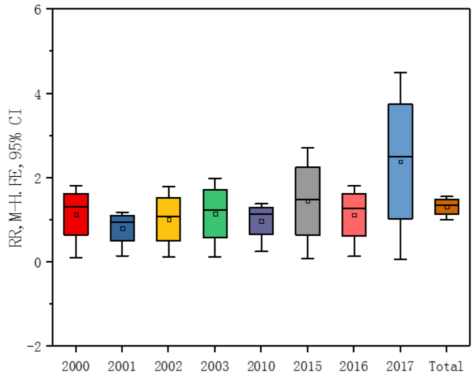

Fig.5 The Box-plot of RR-Fixed

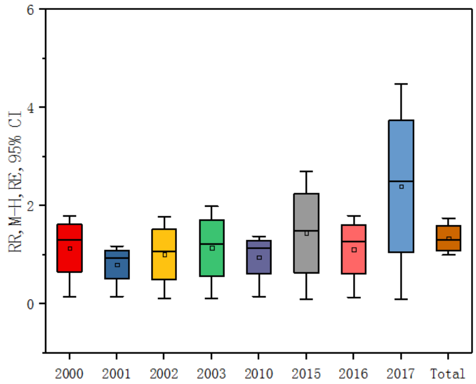

Fig.6 The Box-plot of RR-Random

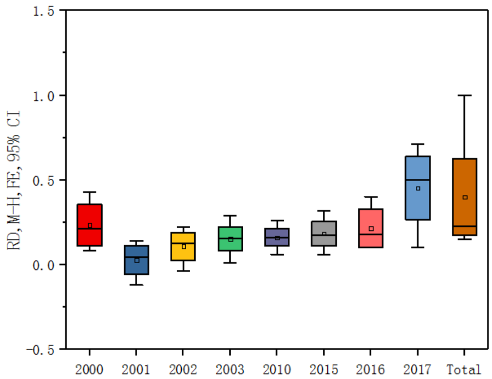

Fig.7 The Box-plot of RD-Fixed

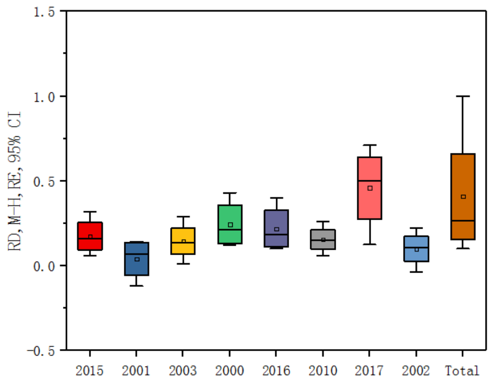

Fig.8 The Box-plot of RD-Random

The experimental results show that the confidence intervals of the effect values in Fig. 3, Fig. 4, Fig. 5, and Fig. 6 are lager, so the results are not accurate. The confidence intervals of the effect values in Fig. 7 and Fig. 8 are smaller. Therefore, RR (Relatve risk) is more suitable for this study than other methods. And the confidence interval of the total effect value in Figure 7 is smaller than that in Figure 8. Overall, the confidence intervals of the total effect values in Figures 3, 5 and 7 are smaller than those in Figures 4, 6 and 8. It shows that the Fixed Effect statistical model is better than Random Effect.

Figures 9-14 are funnel plots of the experiment. It is mainly used for bias analysis. The abscissa is the effect value of the data set, and the ordinate is the standard error of the data set. The smaller the sample size, the more dispersed the distribution is. And the larger the sample size, the more concentrated the distribution is. If there is no bias, it will be symmetrical funnel-shaped. On the contrary, if its symmetry is poor, there is bias.

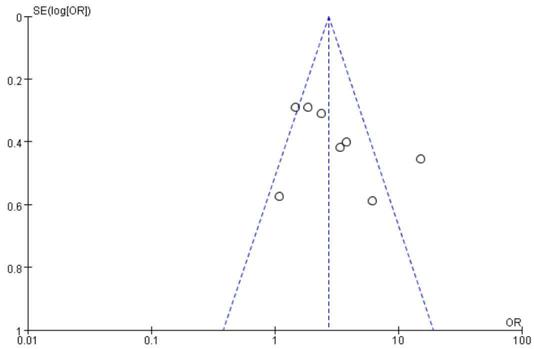

Fig.9 The funnel plot of OR-Fixed

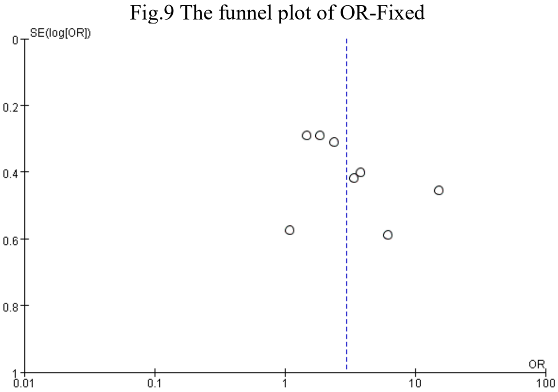

Fig.10 The funnel plot of OR-Random

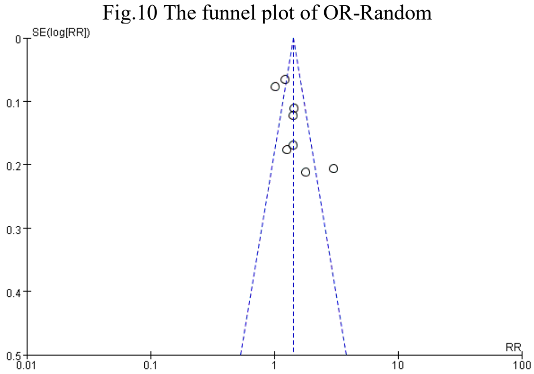

Fig.11 The funnel plot of RR-Fixed

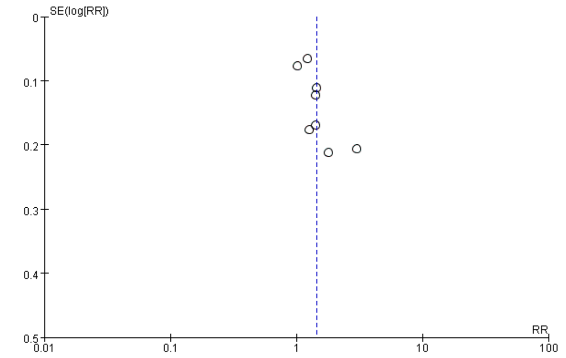

Fig.12 The funnel plot of RR-Random

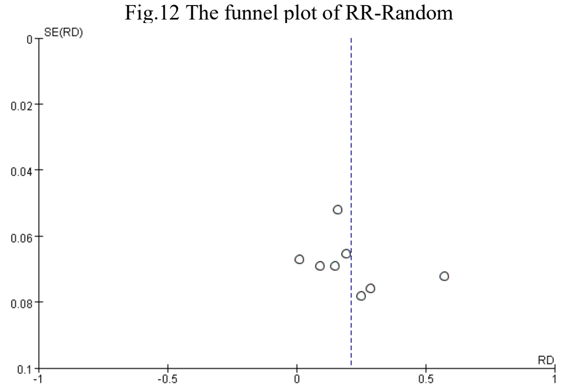

Fig.13 The funnel plot of RD-Fixed

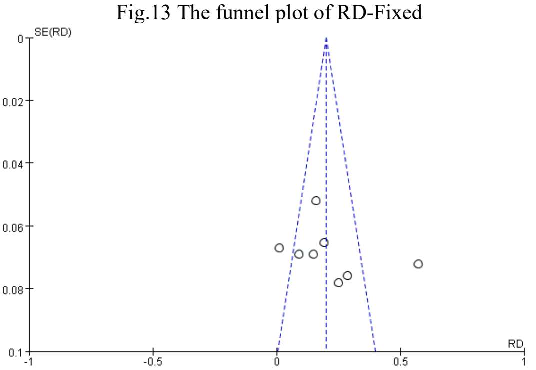

Fig.14 The funnel plot of RD-Random

The experimental results show that the distributions of Fig. 13 and Fig. 14 are more concentrated. And their symmetry is better than others. It shows that the RD (Risk difference) is significantly better than other methods. Overall, the symmetry, centralization and standard errors of Figures 9, 11 and 13 are similar to those of Figures 10, 12 and 14. It shows that the Fixed Effect statistical model is similar to the Random Effect. The analysis of the above experiments show that Risk difference and Fixed Effect are better methods for meta-analysis.

IV. Conclusions

In this paper, we analyses the existing research data of universal intelligence measurement by meta-analysis. It can effectively combine different digital self. And it provides a good research idea for the measurement of digital selves such as human, machine, company, government and institution. The research results of this paper can also promote the development of intelligence science. But this paper has some shortcomings. Because the research quality of different research data is different, and all research data are treated equally in meta-analysis. So, there is some deviation in statistical analysis. We hope that a quantitative standard for research data can be proposed in future studies. This is also our future research focus.

Acknowledgment

This work is supported by the National Key R&D Program of China (2017YFB1400100), and the National Natural Scien ce Foundation of China (61572466), and the Beijing Natural Science Foundation (4162059). Corresponding author is Wen Ji.

References

[1] Turing A M , "Computing Machinery and Intelligence," *Mind (New Series)*, Vol. 59 No. 236, 1950, pp. 433-460.
[2] Cochrane P, "A Measure of Machine Intelligence (Point of View)," *Proceedings of the IEEE*, Vol. 98 No. 9, 2010, pp. 1543-1545.
[3] Hernandez-Orallo J, "Beyond the Turing Test," *Journal of Logic, Language and Information*, Vol. 9 No. 4, 2000, pp. 447-466.
[4] José Hernández-Orallo and Dowe D L, "Measuring universal intelligence: Towards an anytime intelligence test," *Artificial Intelligence*, Vol. 174 No. 18, 2013, pp. 1508-1539.
[5] Prpic J and Shukla P, "Crowd Science: Measurements, Models, and Methods," *Hawaii International Conference on System Sciences. IEEE, 2016*.
[6] Mcgrath M C, "Deviation IQ,", 2011.
[7] Gavin A, Pim C, and Craske M G, "Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis," *Journal of Anxiety Disorders*, Vol. 5 No. 10, 2018, pp. e13196.
[8] Andreas L, Joel L and Barbara M, "Industry sponsorship and research outcome: systematic review with meta-analysis," *Intensive Care Medicine, 2018*.
[9] Legg S and Hutter M, "Universal Intelligence: A Definition of Machine Intelligence," *Minds & Machines*, Vol. 17 No. 4, 2007, pp. 391-444.
[10] Insacabrera J, Dowe D L and Sergio EspañaCubillo, "Comparing Humans and AI Agents," *Artificial General Intelligence*, 2011, pp. 122-132.
[11] Myszkowski N, Celik P and Storme M, "A meta-analysis of the relationship between intelligence and visual "Taste" measures," *Psychology of Aesthetics Creativity & the Arts, 2018*.
[12] Gignac G E and Bates T C, "Brain volume and intelligence: The moderating role of intelligence measurement quality," *Intelligence,* Vol. 64, 2017, pp. 18-29.
[13] Vamsi V and Bose I, "Business intelligence for performance measurement: A case based analysis," *Decision Support Systems, 2018*.
[14] Kahraman C, Sezi Çevik Onar and Oztaysi B, "Fuzzy Collective Intelligence for Performance Measurement in Energy Systems,", 2018.
[15] Legg S and Hutter M, "A Formal Measure of Machine Intelligence,", 2006.
[16] Bien Z, Bang W C and Kim D Y, et al., "Machine intelligence quotient: its measurements and applications," *Fuzzy Sets and Systems*, Vol. 127 No. 1, 2002, pp. 3-16.

[17] Gavane, V, “A Measure of Real-Time Intelligence,” *Journal of Artificial General Intelligence*, Vol. 4 No. 1, 2013, pp. 31-48.
[18] Ke Q and Cheng Y, “Applications of meta-analysis to library and information science research: Content analysis,” *Library & Information Science Research*, Vol. 37 No. 4, 2015, pp. 370-382.
[19] Saxton M L, “Meta-Analysis in Library and Information Science: Method, History, and Recommendations for Reporting Research,” *Library Trends*, Vol. 55 No. 1, 2006, pp. 158-170.
[20] Azucar D, Marengo D and Settanni M, “Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis,” *Personality and Individual Differences*, Vol. 124, 2018, pp. 150-159.
[21] Braga T, GonAlves L C and Basto-Pereira M, et al, “Unraveling the link between maltreatment and juvenile antisocial behavior: A meta-analysis of prospective longitudinal studies,” *Aggression and Violent Behavior*, Vol. 33, 2017, pp. 37-50.