Article

# Reliability of Outputs of Field Random Number Generator Movie Experiments

**Takeshi Shimizu and Masato Ishikawa**

ABSTRACT

Field random-number-generator (RNG) studies have reported that physical RNG outputs can sense field consciousness when many people simultaneously focus on worldwide events or news. The current study focused on the reliability of RNG outputs during an experimental event that evoked audience emotion based on the assumption that the RNG is a measure of field consciousness. Using a split-half method for internal consistency, two previous field RNG experiments (Shimizu & Ishikawa, 2010; 2012) showed positive reliability under an experimental condition ($ICC(1,k) = 0.684$) in which an audience watched a movie that appealed to emotions. Unexpectedly, RNG outputs in the control condition showed negative reliability with a significant correlation coefficient. This suggests that outputs of RNGs have some of the characteristics of quantum effects and have measurement functions in highly focused events. Future tasks are also discussed herein.

**Key Words:** random number generator , random streamer, Rpg102, MMI, FGE, ICC

**NeuroQuantology 2012; 3: 389-393**

## Introduction

Random number generators produce bits (1s or 0s) during real-time processing with an unpredictable mechanism based on a physical random source. In parapsychology, RNGs have been widely used experimentally to investigate the interaction between field consciousness and physical systems. Many studies have shown anomalistic behavior when many people focus on the same events, such as during certain meetings (Nelson *et al.*, 1996), broadcast events (Radin *et al.*, 1996), or the 9/11 terrorist attack (Nelson, 2002; Nelson *et al.*, 2002). Although many types of RNGs with different random sources exist, anomalistic results reported by researchers have essentially showed that RNG outputs were highly biased given a coherent group mind (Radin, 2006) or a focused group energy (FGE; Rowe, 1998) or field. From this point of view, physical RNG devices could be regarded as sensitive and valid "detectors" (Rowe, 1998; Radin, 2006) that measure signals from field consciousness.

In a field RNG study, RNG outputs are basically translated into *z*-scores and *chi*-square statistics. First, based on a binomial distribution, the *z*-scores of the sum of bits in a trial is calculated:

$$z_t = (X_t - n\pi)/\sqrt{n\pi(1-\pi)}, \quad (1)$$

where $n$ is the trial bit size (512 in the current experiments), $\pi$ is 0.50 (meaning that 1s or 0s were observed in equal probablility), and $X_t$ is the sum of bits in a trial. Using the *z*-scores, cumulative *chi*-square statistics were computed:

$$\chi^2_{acm} = \sum_{t=1}^{T} z_t^2, \quad df = T, \quad (2)$$

Corresponding author: Takeshi SHIMIZU
Address: Azuma 4-9-6(301-101), Tsukuba City, JAPAN (code 305-0031) and School of Information and Communication, Institute for Informatics of Consciousness, Meiji University, JAPAN.
Phone: 090-8057-1928
✉ shim1zu@hotmail.com
Received April 19, 2012. Revised July 16, 2012.
Accepted August 2, 2012.
*e*ISSN 1303-5150

where $T$ is the number of trials generated during an event. Note that the *chi*-square value could sense field consciousness.

The current study focused on a novel issue: the question of how reliable measurement achieved by RNG outputs (cumulative *chi*-squares) is. Typical psychology, which involves the development of psychological scales or achievement tests, requires not only confirmation of the validity of questions or test items but also evidence of their reliability for measuring individual differences. In this context, examination of reliability is the first necessary condition for scale development. Studies of RNG outputs, however, differ from typical scale-development studies. First, an RNG was originally no more than a random number generator. Second, parent populations generally provide adequate numbers of participants for scale-development studies. If necessary, items can be added or removed. In contrast, in the case of field RNG, the use of cumulative *chi*-squares as a computation method is basically fixed at this time, and we have not yet determined conditions that may enhance (or decrease) RNG-output measurement reliability.

No test confirming the reliability of RNG outputs is currently available. The purpose of this study was to identify which conditions increase or decrease measurement reliability. If common characteristics of such conditions can be identified, the current study would provide important information for understanding field RNG behavior. The current study considered that RNGs have a certain degree of validity for measuring FGE or the coherent mind of a group and examined methods of evaluating their reliability in a field RNG experiment. If a certain level of reliability could be observed, RNGs may have some function in field consciousness studies.

## Split halves of a single test

Several indices for evaluating the reliability of test scales exist. One of the simplest ways, based on classical test theory, is to use split halves of a single test, which allows evaluation of the internal consistency of items. That is, half of the test items are assigned to Subtest 1 and half to Subtest 2; this is done either randomly or by placing odd-numbered items in one subtest and even-numbered items in the other (Millar, 1995). The correlation or covariance of these halves is computed and used in an equation (several are available) to estimate the reliability of the whole test (Millar, 1995).

This split-half method has an advantage in that it can evaluate reliability based on subsets sampled under identical conditions. To apply this method to current field RNG studies, the entire sequences of the sum of bits from RNG devices were split into two subsets. Cumulative *chi*-squares were separately calcurated as follows:

$$\chi^2_{\mathrm{acm}(Odd)} = \sum_{t=odd}^{T} z_t^2 \, , df = T/2$$

and

$$\chi^2_{\mathrm{acm}(Even)} = \sum_{t=even}^{T} z_t^2 \, , df = T/2. \qquad \textbf{(3)}$$

where all trials were ordered by time and separated based on whether the index of the row number was odd or even. This ensured an equal number of samples for each subset.

After conducting a repeated-measures field RNG experiment, the sample size $N$ of the number of *chi*-square statistics was taken to evaluate the covariance among them. In this framework, the parameter $T$ corresponds to the number of items in cases of psychological scale or test development. The number of repetitions $N$ in a field RNG experiment corresponds to the sample size (number of participants).

## Hypothesis

Theoretically, RNG outputs are essentially independent from one another. Therefore, no correlation between split-half CZs would be observed, and thus reliability would be near zero. However, if the RNG outputs detected signals from field consciousness, two split-half subsets would show a positive correlation during certain coherent events, suggesting anomalistic phenomena. Of course, an issue of RNG quality arises if a positive correlation is always observed, even outside of events. To test such a possibility, outputs under control conditions were compared with those obtained under experimental conditions.

Results of a split method showing clear results were previously reported on a GCP website. The two sets showed quite similar results in the context of data mining. In the current study, the split method was used to evaluate the reliability of measurement by RNG. First, repetitions of field-experiment results are required. Second, this method has an advantage in that it appropriately evaluates results when low cumulative *chi*-squares are observed.

## Method

The current study focused on the results of two previous field RNG experiments (Shimizu & Ishikawa, 2010; 2012). One was conducted at a movie theater in which an audience watched the movie "Departures" (*Okuribito* in Japanese; Shimizu & Ishikawa, 2010). The other was an experiment involving short movies as psychological stimuli (Shimizu & Ishikawa, 2012). Both of the experiments were repeated-measured experiments showing highly emotional movies to an audience. The first experiment was repeated 10 times, and the second experiment was repeated six times. They showed significant biases in cumulative *chi*-square statistics from RNGs despite the fact that the significance was analyzed in total (not individually). No significant biases were shown in the measurements of either experiment. Because these two previous experiments had both control and experimental conditions, we could assess reliability. The control conditions in the first experiment were pre- and post-conditions (before and after the audience saw the movie). The second experiment involved a short, less stimulating movie (a Japanese traditional folk story), which served as the control.

If a RNG detects signals from field consciousness under experimental conditions, it could be predicted that RNG outputs would show some degree of internal consistency. Control conditions, on the other hand, would show no such consistency because no strong signals would be available from field consciousness.

## Hardware and Procedure

In both experiments, Rpg102 hardware was used as a physical RNG device. This device was connected to a PC via a USB port. While the RNG device generated bits in a trial at 512 bits per second, the outputs for a certain time span were recorded by a csv text file in a directory of our notebook PC (VAIO Type-G). Although our software application, developed by Visual Studio C#, to control the device had generated two pseudo-random number generators (PRNGs), the current study did not analyze these PRNG outputs.

## Estimation of RNG Measurement Reliability

Because each condition had a different number of trials (degrees of freedom), all *chi*-square values were standardized by

$$\mathrm{CZ} = (\chi^2_{acm} - df)/\sqrt{2df}, \qquad \textbf{(4)}$$

resulting in $\mathrm{CZ}_{odd}$ and $CZ_{even}$. Based on Rulon's split-half method (1939), reliability ρ was estimated by

$$\rho = 4 \times \mathrm{Cov}\,(CZ_{odd}, CZ_{even})/var(CZ) \qquad \textbf{(5)}$$

where the original CZ variance was

$$\mathrm{var}(CZ) = \mathrm{var}\,(CZ_{odd}) + \mathrm{var}\,(CZ_{even}) + 2 \times \mathrm{Cov}\,(CZ_{odd}, CZ_{even}). \qquad \textbf{(6)}$$

This ρ score was equal to Cronbach's *alpha* coefficient (Millar, 1995).

Although ρ or Cronbach's *alpha* is easy to interpret, the two subsets are not fixed but random factors. Elements within random factors are changeable. Therefore, intraclass correlation coefficients (ICCs; Shrout, & Fleiss, 1979) are more appropriate for use in this study. Although various forms of ICCs exist and, although ICC (3, $k$) is equal to Cronbach's *alpha* (McGraw, & Wong, 1996) or Rulon's ρ in this study, we chose a one-way model ICC (1, k) to compute reliability, where $k$ is the number of independent measurements ($k$ = 2; odd and even) of the same target in a field event. In this one-way model, no systematic error in measurement is taken into account to estimate reliability. Hence, reliability is calculated by

$$ICC\ (1, k) = (BMS - WMS)/BMS, \qquad \textbf{(7)}$$

where $BMS$ is variance (mean square) between targets and $WMS$ is variance within subsets in one way ANOVA model. Total reliability was computed for ρ scores and $ICC$s for all samples under the experimental ($N_e = 28 = 10 + 6 + 6 + 6$) and control ($N_c = 26 = 10 + 10 + 6$) conditions. All calculations for confidence interval (95% CI) and $F$-tests ($F = BMS/WMS$) were performed by SPSS version 19.0.

## Results

Table 1 shows correlation coefficients, ρ-scores computed by Rulon's method, and $ICC\ (1, 2)$ values as indices of reliability under each condition. The ρ -scores and ICCs reflect almost the same information. Additionally, Figure 1 shows the relationship between ICCs and event length. Expectedly, all experimental conditions were associated with positive values. However, all control conditions were unexpectedly associated with negative values. Overall reliability was $ICC\,(1,2) = 0.684\; from\; .323\; to\; .853\; in\; 95\%\; CI$ under the experimental condition, with a significant positive $F$-value ($F(27,28) = 3.16, p = 0.002$). On the other hand, all control conditions showed negative reliabilities. The overall reliability was $ICC\,(1,2) = -1.118\; (from\; -1.374\; to\; -.070\; in\; 95\%\; CI; F(25,26) = 0.42, p = 0.983)$, demonstrating reliability was significantly lower than expected, zero.

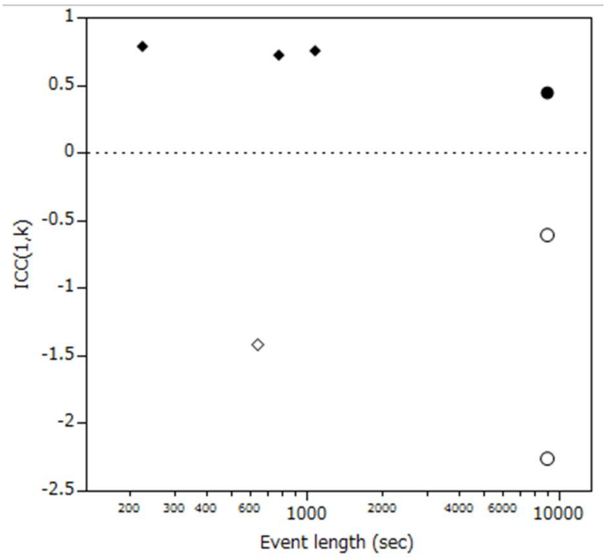


**Figure 1.** Black circles: experimental condition for Okuribito (Shimizu & Ishikawa, 2010), White circles: control conditions for Okuribito (Shimizu & Ishikawa, 2010), Black diamonds: experimental conditions for short movies (Shimizu & Ishikawa, 2012), White diamond: control conditions for short movies (Shimizu & Ishikawa, 2012).

## Discussion

As expected, this study confirmed a positive correlation between split-half subsets under the experimental conditions. However, taking into account that an RNG is just a random number generator, all values greater than 0.60 seem too good to be true because the current results are comparable to those of general psychological scales or tests. In contrast, all control conditions inexplicably showed negative correlations.

These two experiments were originally designed to evoke strong emotions in the audience (Shimizu & Ishikawa, 2010; 2012). Strong emotions or a highly coherent mind in a group could surely enhance the reliability of RNG measurements of field consciousness. Event length or number of repetitions seemed to have limited influence on the results. A rather short movie, 3 minutes in length, showed adequate value. However, the important variable in RNG studies is the number of repetitions, which corresponds to the number of "subjects," and reliability was better when a large number was obtained.

**Table 1.** Correlation matrices and split-half reliability. The ρ by Rulon's split-half method is equal to Cronbach's α or ICC (3, *k*). Parameter *k* = 2 (odd and even) in this study.

| | | | | Reliability | |
|---|---|---|---|---|---|
| | | | | Rulon's *ρ* | ICC(1, *k*) |
| "Okuribito" (Shimizu & Ishikawa, 2010) | | | | | |
| Pre- | Odd | 1.000 | | -1.527 | -2.261 |
| Control (9000 sec) | Even | -0.441 | 1.000 | | |
| *N* = 10 | *SD* | 1.025 | 1.244 | | |
| | *M* | -0.292 | 0.696 | | |
| Film-viewing | Odd | 1.000 | | 0.586 | 0.444 |
| experiment (9000 sec) | Even | 0.415 | 1.000 | | |
| *N* = 10 | *SD* | 0.615 | 0.600 | | |
| | *M* | 0.230 | 0.668 | | |
| Post- | Odd | 1.000 | | -0.599 | -0.609 |
| Control (9000 sec) | Even | -0.232 | 1.000 | | |
| *N* = 10 | *SD* | 0.817 | 0.913 | | |
| | *M* | -0.055 | 0.047 | | |
| Short movie experiment (Shimizu & Ishikawa, 2012) | | | | | |
| No. 1 (Eccentric Puzzle) | Odd | 1.000 | | 0.702 | 0.748 |
| experiment (1080 sec) | Even | 0.810 | 1.000 | | |
| *N* = 6 | *SD* | 1.061 | 1.606 | | |
| | *M* | 0.468 | 0.204 | | |
| No.2 (Sentimental) | Odd | 1.000 | | 0.753 | 0.783 |
| experiment (225 sec) | Even | 0.657 | 1.000 | | |
| *N* = 6 | *SD* | 1.661 | 0.637 | | |
| | *M* | 0.470 | 0.340 | | |
| No. 3 (Comedy) | Odd | 1.000 | | 0.819 | 0.726 |
| experiment (775 sec) | Even | 0.788 | 1.000 | | |
| *N* = 6 | *SD* | 0.901 | 0.538 | | |
| | *M* | 0.095 | 0.574 | | |
| No. 4 (Folk Story) | Odd | 1.000 | | -0.821 | -1.412 |
| control (641 sec) | Even | -0.293 | 1.000 | | |
| *N* = 6 | *SD* | 0.888 | 0.802 | | |
| | *M* | -0.568 | 0.385 | | |

It was found that internal consistency under the experimental conditions was anomalistic. Although previous studies used only *chi*-square tests for results, the index of reliability may be used as an indicator of anomalistic phenomena. This method can evaluate results even when low biased *chi*-squares are frequently observed.

The results obtained under the control conditions were quite unexpected. Notably, the split-half subsets were essentially independent from each other. No reasons for the negative values were found. All control conditions showed significant negative reliability. Additionally, effect sizes (just *r* correlations) were enough to suggest anomalies in RNG outputs.

Negative correlations between subsets of items, where present, indicate that split odd and even trials behaved as if they "knew" each other's behavior and responded accordingly. Consequently, their sum was almost constant. This is similar to a quantum effect. Paradoxically, by adopting a method in which we regard RNG as a measure, we found that a RNG is not a simple measurement with which to detect the existence of an external phenomenon such as field consciousness.

## Future Tasks

Because all of the current analyzed control conditions, including the pre- and post-conditions, were related to the experimental condition with respect to time, the results of the control condition may have been influenced by the adjacent experimental time period. Of course, any premonition made by the RNG would be anomalistic. One future task is to confirm the characteristics of non-event RNG outputs disconnected from experimental conditions. A possible way to accomplish this goal is a bootstrap method or repeated re-samplings from previous RNG outputs that have already been recorded in a database during constant running (*e.g.,* Nelson & Bancel, 2011).

Additionally, the current study analyzed the results of previous experiments using a repeated-measured design. Since current analysis required repeated-measured design, it would be more beneficial if reliability for one (or a few) experimental outputs can be estimated. Fortunately, because the results of the short movie showed high reliability, it is possible that high reliability can be obtained within a dataset; samples were split based on a certain time interval or smaller event unit. Although many field RNG studies assign 1 second to a trial, the current results suggest that field signals are longer than 1 second. These topics will become tasks for future study.

## References

Miller MB. Coefficient alpha: A basic introduction from the perspectives of classical test theory and structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal 1995; 2: 255–273.

McGraw KO and Wong SP. Forming inferences about some intraclass correlation coefficients. Psychological methods 1996; 1: 30–46.

Nelson RD, Bradish GJ, Dobyns YH, Dunne BJ, Jahn RG. Field REG anomalies experiment. Journal of Scientific Exploration 1996; 10: 111–141.

Nelson RD, Jahn RG, Dunne BJ, Dobyns YH, Bradish GJ. Field REG II: Consciousness field effects: Replications and explorations. Journal of Scientific Exploration 1998; 12: 425–454.

Nelson RD. Correlation of global events with REG data: An Internet-based, nonlocal anomalies experiment. Journal of Parapsychology 2001; 65: 247–271.

Nelson RD, Radin DI, Shoup R, Bancel PA. Correlations of continuous random data with major world events. Foundations of Physics Letters 2002; 15: 537-550.

Nelson RD. Coherent consciousness and reduced randomness: Correlations on September 11, 2001. Journal of Scientific Exploration 2002; 16: 549–570.

Nelson RD and Bancel P. Effects of mass consciousness: Changes in random data during global events. Explore 2011; 7: 373-383.

Radin DI, Rebman JM & Cross MP. Anomalous organization of random events by group consciousness: Two exploratory experiments. Journal of Scientific Exploration 1996; 10: 143-168.

Radin DI. Entangled Minds: Extrasensory Experiences in a Quantum Reality. New York: Paraview Pocket Book, Simon & Schuster, 2006.

Rowe WD. Physical measurement of episodes of focused group energy. Journal of Scientific Exploration 1998; 12: 569-581.

Rulon PJ. A simplified procedure for determining the reliability of a test by split halves. Harvard Educational Review 1939; 9: 99-103.

Shimizu T, Ishikawa M. Field RNG data analysis, based on viewing the Japanese movie "Departures" (Okuribito). Journal of Scientific Exploration 2010; 24: 637–654.

Shimizu T, Ishikawa M. Field RNG experiments using short movies: An examination of the focused-attention and emotion hypotheses. Journal of the International Society of Life Information Science, 2012; 30: 17–33.

Shrout PE and Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin 1979; 36: 420–428.