INTERNATIONAL JOURNAL OF RESEARCH & METHODOLOGY IN SOCIAL SCIENCE

VOL. 4, NO. 1 | JAN - MAR, 2018 | QUARTERLY JOURNAL | ISSN 2415-0371

Scientia, Fortitudo & Victoria

QUARTERLY JOURNAL
ISSN 2415-0371

Vol. 4 No. 1

Peer Reviewed Scientific Journal in Social Science

OPEN ACCESS

Berlin Declaration of Open Access to Knowledge in the Science and Humanities
Budapest Open Access Initiatives
The Cape Town Open Education Declaration

ISSN 2415-0371
9 772415 037100 >

IJRMSS

ONLINE & OPEN ACCESS
www.socialsciencepublication.com

BOAD OF EDITORS

INTERNATIONAL JOURNAL OF RESEARCH & METHODOLOGY IN SOCIAL SCIENCE is a peer reviewed scientific journal in social science and its interdisciplinary fields. The journal is published quarterly. The Boards of Editors oversee the acquisition, review, and publication of quarterly research papers in social science. Editorial inquiries may be sent by email directly" Editor@socialsciencepublication.com EditorIJRMSS@gmail.com.

Vol. 4, No. 1, p.3 (Jan. – Mar. 2018). ISSN 2415-0371 (online)

*Candidate.

TABLE OF CONTENT

Language Production
Part 2

Dante Roberto Salatino★

About the author
★Dante Roberto Salatino is a researcher of the Institute of Philosophy and of the Institute of Linguistics - Lecturer in the General Psychology Department – Lecturer in the Philosophical Aspects of Physical-Mathematical Science Department - Faculty of Philosophy and Letters - Teacher and Researcher in Artificial Intelligence in the Mechatronics Career - Faculty of Engineering - National University of Cuyo - Email for correspondence: dantesalatino@gmail.com

ABSTRACT
In Part I, we have analyzed various aspects that condition the production of conventional language. For this purpose, we have described the biological elements that intervene the production of the word and then stop at the psychological aspects of production, something we did through an approach to one of the most critical pragma-linguistic manifestations: '(dis)politeness.' In this Part II, we will analyze one of the most known production models of those proposed by Psycholinguistics, which, that considers the processes involved in the production of language are of the psychological/cognitive type, something that also supports Cognitive Psychology. We will conclude this study with the proposal of a model of conventional language production from their 'junction' with natural language, that is, how one 'transforms' into the other; a transformation that is possible given series of bio-psychic processes and mechanisms.

Keyword: Conventional language, natural language, models of language production, Psycholinguistics, Transcurssive Logic.

CITATION:
Salatino, D. R. (2018). "Language production – Part 2" *Inter. J. Res. Methodol. Soc. Sci.*, Vol., 4, No. 1: pp. 5–17. (Jan. – Mar. 2018); ISSN: 2415-0371.

1.0 INTRODUCTION

The linguistic production concentrates the concurrence of several proposals. Any of the approaches show a typical pattern that has been proposed since the last quarter of the previous century on. For this reason, we will base our analysis explaining the aspects that all the proposed models 'solve' in a conventional manner and with this I mean, based on the empirical evidence and contrasting it with some *ad hoc* computational model.

Psychology with cognitive orientation and psycholinguistics consider that the processes that intervene in the production of language are of three types: psychological/cognitive, linguistic/grammatical and communicative/instrumental (Belinchón, 1996, p. 536)

In this Part II, we will deal primarily with the first. Based on the assumption that the use of language implies, on the part of the speaking subject, that a selection of the content of their messages from a series of representations that would be stored in memory accompanied by general and specific processes that would involve both their attention, and their voluntary motivation. At the same time, this selection would include the realization of a not very precise series of 'mental operations' on these representations.

2.0 PRODUCTION MODELS

To address the psychological/cognitive processes that would intervene in the production of language, we will take as a point of reference the cognitive model of Willem J. M. Levelt. (1999)

This model proposes that three phases or stages of processing would be involved in the task of language production, namely: 1) A planning or conceptualization phase: which would allow the selection of the content of the message to communicate. As a result of this stage, we would obtain the prelinguistic message that is the product of a gadget called a conceptualizer. 2) A phase of linguistic coding or formulation: in which translate the prelinguistic message into the linguistic format, which implies the progressive (incremental) specification of structural units that will intervene in the speech to configure a phonetic plan or ordered representation of units linguistics that make up a sentence. The result of this stage is the phonetic plan or internal speech; and 3) Peripheral or articulation processes: using which the phonetic plan is translated into a code or motor plan that triggers the sequence of muscular movements that produce speech.

We will analyze this model in detail and to do so we will base ourselves on the work of Levelt et al. : ‘A theory of lexical access in speech production’ of 1999.

In this work, the authors propose a series of mechanisms involved in the linguistic production and suggest, also, some of the supposed cognitive processes that underlie this production, which we will now consider.

In a particular ontogenetic introduction, the work describes how an infant manages to produce its first significant words. Acquiring, on the one hand, the notions of acting, interacting, the causal and temporal structure of events and the location/permanence of objects which would give the child the ability to create the first lexical concepts indicated by a verbal label. This label would initially be exclusively auditory patterns taken from the environment.

On the other hand, a series of bubbles appear with their articulatory gestures, which, by repetition and concatenation, increasingly resemble the real expression they are listening. In this way, it is constituted a repository of speech patterns, although still devoid of meaning.

According to this proposal, the real word begins when the child manages to connect a particular babble to a specific lexical concept, that is when it occurs the coupling of a conceptual system and a motor articulation system. This system grows until a phoneme process separates the gestures of the words, converting the latter into a concatenation of phonetic segments.

Phonetic coding would become a phonological coding system, origin of the adult, phonemic and metric patterns, respectively. The more abstract representation guided by phonetic coding would create the appropriate articulatory gestures.

The conceptual root, in the first instance, allows the creation of sentences of several words, whose order is only semantic, that is, where relations established between lexical concepts prevail. As the authors see it, a precise genetic mechanism drives a restructuring of the conceptual system, giving rise to a process of syntactization.

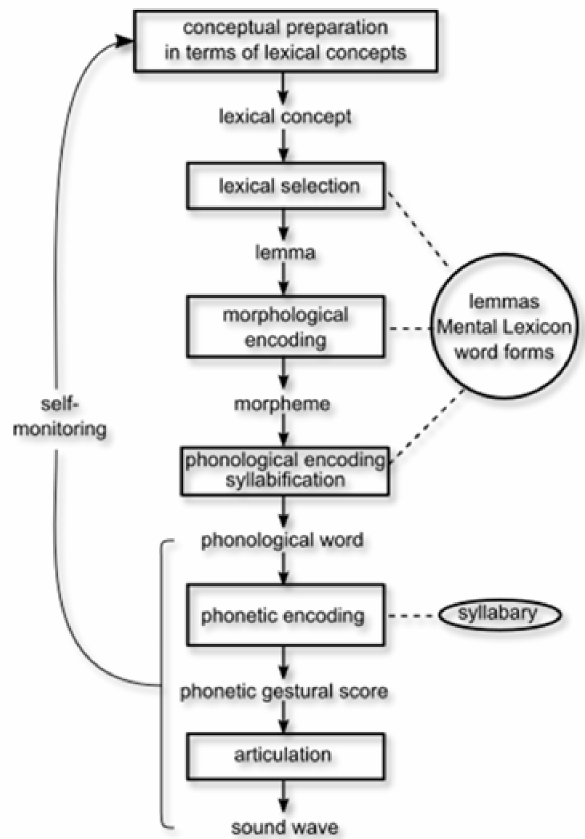
All of the above allows lexical concepts to acquire a specific syntactic category. Thus, the verbs acquire the specifications of the way in which the semantic arguments are projected on the syntactic relations; nouns acquire their properties of syntactic correspondence, etc. Technically speaking, the child develops a system of slogans, that is, a package of syntactic information for each lexical concept.

At the same time, the child quickly acquires a relatively small set of frequently used functional words, which mostly serve syntactic functions given the poor relationship they have with lexical concepts. This system of *lemmas* constitutes a large part of the operation of the system at the age of four years. From then on, the production of a word always implies the selection of the appropriate *lemma*.

Finally, the original double system becomes a four-level processing device: 1) activation of lexical concepts, 2) selection of *lemmas*, 3) morphological and phonological coding of a word in its prosodic context and 4) phonetic coding of the word.

In Figure 1 we can see the different operational levels proposed by this theory and their respective products. But also, in a very succinct way, are outlined those cognitive processes that are supposed to underlie this mechanism and that determine the understanding of the language.

Figure 1: Theoretical delineation of word production in parallel with the self-monitoring of the normal mechanism of speech comprehension. (extracted and modified from Levelt, 1999)



Before analyzing in any detail the different levels, I want to dwell on two terms that have a special psycholinguistic relevance; I refer to the concept of the *lemma* on which an essential part of the theory of language acquisition and *mental lexicon* is based, the common factor to all cognitive approaches.

2.1. Lemmas

The concept of lemma was introduced by Kempen and Hoenkamp in 1987, within the framework of the proposal of a sentence construction mechanism, which they called incremental procedural grammar (IPG), to which they assign a specific psychological and linguistic plausibility. According to these authors, a speaker can construct a syntactically coherent statement from a series of syntactic fragments that contribute with their meaning to the general purpose of enunciation. This increase in significance, from left to right, as a way of producing sentences, is the nuclear capacity of the presented incremental grammar.

The psychological plausibility of this artifact is based on the fact that this grammar seems compatible, at a satisfactory level of linguistic plausibility, as accepted by current theories of grammar, with this way of structurally forming judgments.

The essential design characteristic of IPG, which gives rise to its psychological and linguistic 'properties', is the concept 'Procedure + Stack.' Phrases are constructed, not by a central processing, but by a set of syntactic procedures (modules) that work in parallel, in small parts of the sentence, so they have only a limited vision of being their only means of communication, a battery.

The IPG includes object-complement constructions, interrogatives and the order of the words in the main and subordinate clauses. It deals with dependencies and coherence between grammatical elements, as well as self-correction and elliptic responses to questions.

The IPG was implemented as a generator of incremental sentences in Dutch, written in LISP, the second oldest high-level computer programming language and the basis for the first language used in artificial intelligence. LISP comes from the acronym: LISt Processing. Chained lists are one of the critical structures in this language, and in fact, their source code itself is composed of lists. This language was also used in artificial intelligence in its native form.

Next, the authors describe the cognitive processes that underlie the production of a sentence, saying that they are usually grouped under the categories: content, form, and sound. They then state that a group of activities is devoted to the conceptual (semantic) planning of the content of the expressions of the language. A series of conceptual structures that are 'understandable, interesting and non-redundant' for the listener are selected to be verbalized. This linear conceptual structure can be divided into a sequence of messages, where each can express a complete or partial sentence.

These and other annexed activities can be called *conceptualization*. The second group of processes is responsible for translating the meaningful content in the form of a phrase, a process that they call *formulation*. Finally, the syntactic and morphological structures constructed by the *formulator system* are passed on to the *mechanisms of expression* for their final articulation.

In this apparatus designed to construct sentences that express an intention of a supposed speaker, the mechanisms of the lexical-syntactic stage are privileged. These tools have to do with a syntactic processor, a device widely used in cognitive linguistics and psycholinguistics, which can operate both on the format of grammar rules, and on the structure-function, hence the appellation of procedural.

The authors adduce that all the procedural grammars that existed until the appearance of their article had been developed in ‘non-psychological contexts.’ It is unlike their proposal that when contemplating a ‘computational simultaneity’ to be treated in parallel, through active procedures or modules the different branches of a syntactic tree, such as those taught by Chomsky (1957). Then, each process is constituted, in an ‘expert’ specialized in assembling different types of syntactic elements, admitting in this, a high degree of psychological plausibility.

Without going into much detail, two types of syntactic procedures are distinguished: categorical and functional. The categories, in turn, differ in two varieties: framing and lexical. With this base, a system of lexicalization has the mission to inspect the conceptual structures to look for the words or expressions in the mental lexicon, that represent the intention of the speaker. This lexicalizador is the one that initiates the process of formation of the trees. From here, the 'conceptual structures' are only used for inflection calculations or the insertion of function words.

The process of lexicalization consists, definitively, in recovering 'lexical entries' that are procedural, that is, they include a list of one or more calls to specific syntactic processes. The authors denote such entries with the term: lemma (form), to distinguish them from lexemes (lexical meaning).

From here the 'psychological aspects' managed by this computer program are detailed, among which are: conceptual evolution achieved through an iteration process, treatment of interrupted syntactic units (repair) and ellipsis, speech errors, etc.

Finally, we must pay attention to the fact that, from the IPG, the syntactic analysis (syntactic parsing), as part of the process of perception of language, is markedly similar to the syntactic formulation, as part of the process of language production.

Because it is relevant when highlighting the firm adherence of all these proposals to the computational metaphor, I will list such coincidences: (1) The lexicon manages both the analysis and the syntactic formulation, that is, they operate by syntactic information stored with the individual words in the lexicon. (2) Both processes use this information to build a syntactic tree, with those words as terminal elements and (3) Both have facilities to 'grow' syntactic trees from left to right. The analyzer needs them to add new words to the syntactic tree analyzed, the formulator to compute next (incremental production). The origin of the words is different: they come from the speech recognition in the case of the analyzer and the lexicalization in the case of the formulator.

2.2 Mental Lexicon

Almost an as something natural, it is assumed that words cannot be randomly stacked in mind and this for two obvious reasons: first, because of the significant amount of them and secondly, how could we find them in a second fraction? The cognitive psychologists accept this kind of Vulgate who has 'demonstrated' that human memory is flexible and extensible, as long as the information is structured, which supposedly proves the existence of a highly organized mental lexicon (or dictionary), much bigger and more complex than an ordinary dictionary.

The psycholinguistics has few tools to build a model of the mental lexicon, other than the intelligence of the researcher and a heterogeneous collection of clues. We can identify at least four: (a) the search for lost words and the *lapsus linguae* of the regular speakers, (b) the effort of the search of words in people with speech disorders, (c) psycholinguistic experiments and (d) the findings of theoretical linguistics.

2.2.1 Lost Words

The importance assigned to this clue lies in a presumed neighborhood between the words, either because they are 'stored' near or because being far from each other, there are high bonds that bind them, as Freud suggested in "Psychopathology of everyday life" of 1901. The information provided by this clue needs more convincing evidence.

2.2.1.1 Lapse Linguae

This involuntary error in spontaneous speech has been considered as one of how are revealed a series of speech mechanisms that remain hidden.

Lapsus can be classified into two types: assembly errors and selection errors. Of the former, it is considered that they contribute little on the structure of the lexicon, but not the latter, which suggests that the error occurs when selecting a term from the 'store of mental words.'

This is privilege selection errors because it is assumed that if one makes this type of errors is because he accidentally selects a word that is 'close' to the word sought or 'white.' Freud thought that these mistakes revealed thoughts involuntarily suppressed, making their way to the surface; psycholinguists, however, consider that they are less secret thoughts than information about the mental lexicon.

Selection errors can occur between similar meanings, sound similarity or both. The data provided by this track are not categorical, much less, since it is complicated to distinguish if it was produced by having wrongly chosen the word in question or by ignoring it directly and it cannot be ruled out that the speaker is suffering from a disorder of the speaks.

> "The paraphasia (misuse of words) observed in aphasic patients [conduction aphasias] does not differ from the misuse and distortion of words that normal people can see in themselves in a state of fatigue or division of consciousness. attention or under the influence of disturbing emotions." (Freud, 1891, p. 29).

2.2.2 Without Words

Aphasia is a severe speech difficulty that usually arises as a result of a stroke, severe brain trauma or a specific psychopathology. Psycholinguists study aphasias because they assume that the appearance of their symptoms would not be possible if the healthy cognitive system were not organized in a certain way. The existence of patients who show selective alterations such as, for example, remembering verbs but not names, could indicate, they argue, that these syntactic elements are organized in different subsystems of the lexicon.

There are two crucial drawbacks with aphasics. The first and most obvious is that the damaged brain or psyche may not be representative of its normal counterparts. The second and not least has to do with the wide variety of underlying causes, which may conspire so that the same evidence can be produced for different reasons.

2.2.3 Psycholinguistic Experiments

Derived from both experimental psychology and the field of theoretical linguistics, these experiments, whether based on the measurement of time or the investigation of a particular combination of words, suffer from almost insurmountable difficulties. Such as the fact of having that to pose a very reduced and highly improbable situation to be able to have an absolute control of the considered variables or to privilege the manipulation of combinations of words to obtain a meaning, instead of the words themselves.

All the elements considered previously make the task of elaborating a mental lexicon model difficult. So it ends up proposing 'mental maps' based metaphorically in a library or a computer, where are administered by convenient subdivisions. In a word, from its syntactic aspects to the pragmatic circumstances that concern it, passing through its possible meanings, absolutely biasing, in this way, the implication that this type of constructions may have on the psychic structure/function.

2.2.4 Findings of the Theoretical Linguistics

Returning to the stages contemplated by Levelt in the process of producing words, we see that characterizes them from conceptual preparation to the initiation of the articulation, pointing out in each one, their output representations. That is to say: lexical concepts, lemmas, morphemes, phonological words and the phonetic-gestural results that are executed during the articulation, although without resolving if these stages overlap in time or if they are strictly sequential.

Without being exhaustive, we will give some details of each of the stages:

- *Conceptual preparation*: refers to the different forms of activation of the lexical concept of a word. This activation is justified assuming that, like all open words and most of the closed words are significant, the intentional production of a meaningful word "always" involves this conceptual activation. That sometimes includes a problem of verbalization, that is, discerning the communicative intention from a message that not only represents a concept but a real conceptual structure. This stage of the theory was modeled by a conceptual network, which consists of a series of linked 'concept nodes,' which are conveniently labeled to indicate the character of the link.

- *Lexical selection*: allows to retrieve a word or better a lemma from the mental lexicon, before the expression of a lexical concept. The model achieves this by linking a layer of lemma nodes to the conceptual network, assigning a lemma node to each lexical concept. An 'activated lexical concept,' disseminates its activation to its slogan nodes through a 'statistical mechanism' that favors the selection of the most activated slogans. Although this is the most important selection mechanism, in the case of function words, it is driven by the syntax and not by its conceptual value. In this way, selecting from the slogans creates a suitable syntactic environment for the chosen word. This mechanism is intended to simulate also something similar to what Slobin (1987) called 'Thinking for Speaking,' that is, to achieve the necessary grammatical agreement and the temporal relevance of an expression from the conceptual representation.

- *Morphological and syllabic coding*: after having selected the word syntactic or lemma, the speaker is about to cross the boundary that separates the conceptual/syntactic domain from the phonological/articulatory. This task consists in preparing the appropriate articulatory gestures for the selected word, in its prosodic context. First, it proceeds to recover the phonological form of the word in question, from the mental lexicon. According to the theory, access to the phonological structure involves activating three types of information: (a) the morphological distribution, (b) the metric form and (c) it's segmental distribution, leading to the syllabification process, making it not stored in the lexicon mental. This process was modeled assuming that the segments of a morpheme or phoneme are available at the same time, linked using appropriate links to indicate their correct order.

- *Phonetic coding*: this theory gives a partial solution to this aspect. The general objective is to indicate how the gestural results of a phonological word are computed. In the computational model that supports the theory, this coding was simulated by activating the syllabic effects. Derived from the segments of the phonological syllables, assuming that the speaker must have direct access to frequently used results, thus ensuring greater coherence intersyllabic In this way, it is assured that the articulation of the phonological word begins as soon as all of its syllabic results have been recovered. It is recognized that this procedure does not cover all alternatives.

- *Articulation*: the gestural results of the phonological word are finally executed by the articulatory system, although the operation of this system exceeds this theory.

- *Self-monitoring*: the first thing we hear, when speaking, is our voice, so we can control the appearance of problems in our speech, in addition to our interlocutor, which implies the normality of our perceptual system. Although the theory does not go beyond the initiation of the articulation, it does suggest a management of the self-correction of the 'internal speech,' that is, it can control some internal representations that occur during the coding of speech. Levelt proposes that 'internal speech' is a representation of phonetic coding, therefore, accessible from the analysis of the syllable.

After making a detailed analysis of the stages involved in the production of the word, the authors make a precise detail of the computational model that allowed to 'corroborate' the theory. On which, we have highlighted some of its characteristics when describing each one of the stages, so we are not going to abound in details. Suffice it to say that it is based on a network that shows a propagation of its forward activation, that is, no is applied an output of a node to the input of a preceding node, with the possibility of inhibiting certain nodes as needed. The computational model achieved was baptized as "WEAVER" and is ingeniously programmed so that, using the contribution of a series of suitably structured data, it complies with the different stages proposed by the theory. Adjust the data provided according to the results of various psycholinguistic experiments, such as, for example, the 'asynchronous start stimulus,' in the analysis of the influence of distractors on semantic facilitation and inhibition, to demarcate conceptual interferences or possible interactions between semantic and orthographic factors.

Finally, the authors offer two types of evidence that in some way try to prove the plausibility of the proposed theory. This evidence is those referred, on the one hand, to the reproduction with enough approximation of the empirical times obtained in experiments that suppose an exploration of each one of the suggested stages. On the other hand, those based on capturing functional brain images, such as Event-Related Potentials and Magnetoencephalography, simultaneously with the operation of a detailed model of cognitive processing with data obtained from the experimental task.

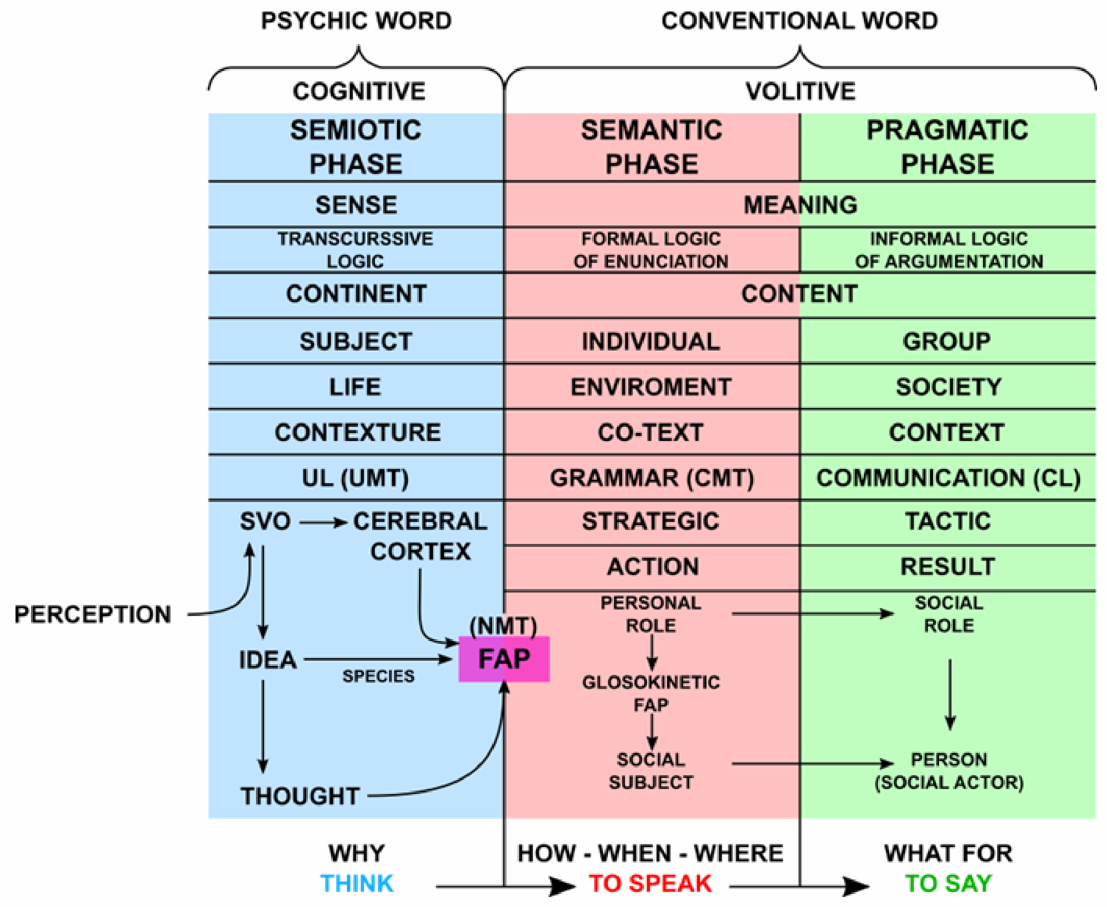
By way of summary, we can say that this theory proposes, on the one hand, a system governed by concepts whose objective is to select those words (*lemma*) from the 'mental lexicon' that more adequately express specific intention of the speaker. On the other hand, a system that prepares the articulatory gestures for the selected words in the context of the enunciation. Both systems are linked in a somewhat lax way. It provides a computational model that simulates each of the stages proposed in theory, which yields results that are empirically tested, which, according to the authors, shows that the theory reaches a reasonable level of congruence and plausibility.

3.0 PROPOSAL OF A PRODUCTION MODEL

Although, as we already anticipated, unlike conventional language, there is no production as such of natural human language. At this section and a counterpart to the model previously analyzed we will suggest a model of conventional language production at the point of departure of the 'junction' with natural language, that is, how one is 'transformed' into the other.

From the beginning, I want to highlight, on the one hand, the psychobiological nature of this connection and on the other, that it is represented by a fixed action pattern (FAP), see Appendix. Figure 2 shows an overview of the model.

Figure 2: A Production Model
References: UL: universal language - UMT: universal maternal tongue; CMT: conventional maternal tongue - CL = everyday language; SVO: subjectivon - NMT: natural maternal tongue - FAP = fixed action pattern.



We can observe in the previous figure that the proposed model covers two well-differentiated aspects. On the one hand, it deals with the domain of the word psychic, that is, the cognitive and therefore of a mostly unconscious nature; and on the other, with the realm of the ordinary word, where the conscious will prevails.

On the other hand, the model can be divided into three specific phases: (1) *Semiotic phase* or sense, which covers the cognitive domain in its entirety. The volitional domain, or meaning, in turn, is divided into: (2) *Semantic phase* or that of the literal meaning or that of which it is spoken; and (3) *Pragmatic phase* or the real meaning or what we want to say.

The semiotic phase is conditioned by the transcurssive logic, while the semantic phase is conditioned by the formal logic of the enunciation; finally, the pragmatic phase conforms to the designs of the informal logic of argumentation. These logical ties make the semiotic phase act as a continent of the other two, which constitute their occasional content in the face of a specific communicative situation.

The only resident in the semiotic phase is the subject. The semantic phase is inhabited by the individual and the pragmatic phase by the group. Therefore, the objective to be controlled by the semiotic phase is life (biological, psychic and social); the semantic phase is in charge of the relations of the individual with the environment, remaining for the pragmatic phase, the control of the relationships with the others, that is, with society.

The semiotic phase is contexture dependent. However, the semantic phase is dependent on the co-text or linguistic context, while the pragmatic phase is dependent context.

The universal maternal tongue, the one that derives from the universal language, is the rector of the semiotic phase. Grammar, represented by the conventional maternal tongue, dictates the rules of 'good speech' that institutionalize the semantic phase and the communicational norms, put into evidence by everyday language, in its various genres, governing the pragmatic phase.

The model shows how the word is elaborated (semiotic phase) and once elaborated, how it is projected in a strategic action (semantic phase) to obtain a tactical result. In other words, what needs to be done to live together, that is, to stay alive socially through the recognition of others. Although it depends on our behavior, it is strongly conditioned by how we communicate with the other, something that as we saw, it is wholly expressed, for example, in politeness.

The specific mechanism proposed in this model begins with the perception of an act of emission made by another, to which we have to answer, or the perception of a communicative situation that requires our intervention as speakers. This perception that is restricted by the subjectivon (See Appendix) inherited from our mother. That is, by its way of ‘seeing’ the subjective reality, that generates, in the first place, a psychic structure according to and adjusted to chronological time, represented by an idea, and second, it gives rise to the constitution of an essential cognitive structure: the species (Salatino, 2014, p. 93).

The species, under the strict surveillance of the universal maternal tongue, represented by the subjectivon, allows the elaboration of the natural maternal tongue. On the other hand, the idea, the structural, is functionalized giving rise to thought, which by adjusting to what is established by the natural maternal tongue, converges on the cortical structures, to leave created the refined mechanism that will allow preparing everything necessary for the emission of the word. This mechanism, once adequate to the desired emission, is semi-automated so that it can be used when similar words have to be used regarding their meaning, that which we acquired by imitating our elders. We have called this mechanism Glosokinetic FAP.

The name of glossokinetic assigned to the FAP that serves as a link between the natural maternal tongue and the conventional maternal tongue is quasi-inappropriate since the etymological meaning refers to the movements of the tongue and the truth is that less than 10% of the communication is words. Almost 40% is transmitted through the inflections of the voice, and around 50%, through body language (Mehrabian, 1971) which includes: postures, gestures, movements of the hands, eyes and even movements respiratory. For the above reasons, here, the term glossokinetic consists of all of the above and thus may represent the link of the social subject (the individual) within its personal role, which indicates: how, when and where to speak, fulfilling the function of the semantic phase of this model.

Finally, entering the pragmatic phase, the person or social actor adopting its social role, adapts its communication, according to a particular intention that has as a purpose to achieve a specific effect on its interlocutor, that is, a justification for what to say.

4.0 CONCLUSION

The first criticism that should be made to the Levelt model is that to demonstrate that the speaker is a highly sophisticated information processor. That is capable of transforming intentions, thoughts, feelings, into articulate fluid speech, it uses a real information processor to which it assigns as a fundamental task, to execute all the cognitive processes that are involved in the production of speech, as proposed by the theoretical model.

A second criticism has to do with the means used to corroborate the plausibility of the model. On the one hand, to adjust the model are used empirical data obtained in experiments designed on restrictive basal conditions and very oriented to what is intended to achieve. Then, using a series of computer gadgets, it manages to produce results that adjust quite well to the experimental results, especially those that have to do with the times involved in each of the supposed stages through which the production process has to pass. On the other hand, the use as a method of recording the cognitive activity of a variety of evoked potentials called Event-Related Potentials (ERP) which according to the authors allows them to corroborate in live performance, three essential aspects of the model. Such as 1) studying the temporal sequence of the different phases of the process, as well as analyzing the separate 'time windows' within each stage that occurs in the preparation of the various responses. 2) obtain the spatiotemporal course of cerebral cortical activation during lexical coding and thus compare it with the temporal stratification proposed by the model and 3) be able to study an isolated process. The above is based on suitably designed experiments that showed that could isolate a variable that affects only one component of the process, for example, in this theory, the variable that controls the frequency of the words, just affects the duration of the morphological coding. Then, any concomitant variation in the space-time course of the brain activation 'must' correspond to the functioning of only one of the components of the processing.

The contributions of the ERP, from the medical point of view, are still under discussion, including those of Cognitive Evoked Potentials, such as P300, used to analyze tasks involving intentional discrimination, and which is accepted as a valid index of processing of central information engaged during decision making related to a given function.

The discussion about the validity assigned to these methods arises from the lack of their neural correlates. So, they are a standard option for psychological tests and even for the measurement of cognitive function, merely because they are easily reproducible, but not because they allow adequate temporal-spatial discrimination of cortical functions. Its lack of perception is because it is an aggregate record (a post-potential) that arises from a large number of neurons and that supposedly implies, when evoked, some form of consciously making a situation or making a decision. The anatomical substrates of the P300 have not been clarified, nor the origin of their component waves, nor even how are generated the evoked potentials, that is if they are induced by the incoming stimulus or produced by the reaction of a neuronal population that focuses its firing patterns based on a stimulus. On the other hand, it is not possible to discriminate the influence of contextual information, so the results are further diluted. In conclusion, the ERPs are a sensitive but very unspecific tool.

Finally, a third criticism derives from two core issues when dealing with aspects related to cognition. I refer, on the one hand, to the use of semantic projections originated in the syntax as if they were valid concepts, if not, definitions taken from a mental lexicon, in replacement of real meanings. And on the other hand and much more critical, to the arbitrary configuration of the so-called cognitive processes, such as subroutines of a computer program (WEAVER) where the solution is given, only by finding the right algorithm so that this routine can achieve a previously chosen result; thus, transforming itself into the description, not of the complex mechanisms that the psyche implements to produce speech, but of the twists and turns that the systems analyst and the computer programmer have to overcome to solve the problems that the programming language elected to develop the model.

This last difficulty is unavoidable because usually, the models used to endorse the theoretical proposals, are simulators and not emulators. Simulating means obtaining results similar to those sought, that is, imitating the operation of the studied system, without taking into account the underlying processes that produce it. To emulate, on the other hand, is to obtain results as a consequence of accurately modeling the functioning of a series of coordinated processes that seek to mimic natural processes. When in an emulation, the results thrown by the model are similar to the empirical data, it is then legitimate to extrapolate, and be able to say that 'possibly' the mechanisms involved in the natural process resemble those proposed; Doing this with a simulation is not correct. The production model presented by us is closer to emulation than a simulation.

The provenance of the presented model is corroborated, beyond its solid neurobiological basis, if we bear in mind that we can support Plato's suggestion. It asserts in other terms, that it is not the same to have something to say, that to have to say something. With this model, it is possible both. To define the 'have something to say' as a product of a strategic-tactical integration of the said, like the unfounded saying, which arises spontaneously without any intentionality, and all this without having to appeal to implications or ‘implicatures’ that lack support in subjective reality.

APPENDIX

GLOSSARY

Conventional Maternal Tongue (CMT): is the socio-cultural support of the language that is acquired by imitation, after 18 months of life (everyday language). Represents 'how is the subjective reality' from the social point of view.

Fixed action pattern: (FAP) (Llinás, 2003) Behavioral expressions that arise from the operation of modules of motor activity and that show, both aspects of an automatic nature that have a strategic purpose, as well as tactical aspects of a voluntary nature. In this work they are considered the basis of the generation of habits and they are divided into three types: a) innate, b) modifiable by experience and c) acquired. Each of these categories is operated at different levels of brain complexity according to phylogenesis of the nervous system.

Natural Maternal Tongue (NMT): is the biological support of language acquired in the first 18 months of life (natural human language). Represents 'how is seen the subjective reality' from the biological point of view.

Subjectivon (universal linguistic pattern): is in each of the domains in which is structured the subjective reality and represents the basis of the different ways of 'seeing reality.' His biological equivalent is in the mitochondria, and like it, has as we saw a circular double-stranded DNA structure on two complementary strands or chains, a heavy or external (the nucleoid) and one lightweight or internal (the gamete). It also has some organelles that are designed to 'sustain' aspects of the universal language, and of the future natural language, screened in the conventional language, once acquired. The possible nucleoids of a subjectivon are just 6, and represent all combinations without repetition of the subject (S), the object (O), and the change (V). That is dextrorotatory variants: SVO, OSV, VOS; and the levorotatory variants: OVS, VSO, SOV.

Universal Maternal Tongue (UMT): is the mode of behavior of a given subjectivon. The tendency to externalize a specific form becomes manifest in the order of the elements in its nucleus. Represents 'how is seen the subjective reality' from the psychical point of view.

REFERENCES

Aitchison, J. (1990). *Words in the Mind - An Introduction to the MentalLexicon* - Oxford, Basil Blackwell, Inc.

Belinchón, M. et al. (1996). *Psicología del lenguaje - Investigación y teoría* -Valladolid, Simancas Ediciones.

Borges, J. L. (2011). Borges – *Obras Completas* – Tomo Diez – Buenos Aires, Editorial Sudamericana.

Breznitz, Z. (2008*). Brain Research in Language* - Haifa, Springer.

Freud, S. (1901). *Psicopatología de la vida cotidiana* - Volumen VI - Buenos
Aires, Amorrortu Editores.

Freud, S. (1973). *La afasia* (Trad. A. Alcalde) - Buenos Aires, Nueva Visión.

Kempen, G.; Hoenkamp, E. (1987). *An Incremental Procedural Grammar for Sentence Formulation* - Cognitive Science, 11, pp. 201-258.

Levelt, W. J. M. et al. (1999). *A theory of lexical access in speech production.* Behavioral and Brain Sciences, 22, pp. 1-75.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique* -California, MIT Press.

Mehrabian, A. (1971). *Silent Messages*. Belmont, California, Wadsworth Publishing Company, Inc.

Ramírez, S. F.; Hernández, J. F. (2011). *Potenciales evocados relacionados con eventos cognoscitivos* - Cap. 20 in Guía Neurológica de la Asociación Colombiana de Neurología - In http://www.acnweb.org - 05/12/2017.

Salatino, D. R. (2014). *Psyche – Structure and Function* – Mendoza, Argentina, First Desktop Publishing. ISBN: 978-987-33-5702-2.

Slobin, D. I. (1987). *Thinking for Speaking*. Proceeding of the Thirteenth Annual) Meeting of the Berkeley Linguistics Society, pp. 435-445.

Financing Preference for MSMEs in Rural Tanzania

Lubawa, Galinoma★
Shirima, Andrew★★
& Nandonde, Felix Adamu★★★

About the author
★Lubawa, Galinoma, Institute of Rural Development Planning (IRDP); Tanzania,Email: galinoma@gmail.com
★★Shirima, Andrew, University of Iringa(UoI) – Tanzania
★★★Nandonde, Felix Adamu , Sokoine University of Agriculture(SUA) - Tanzania

ABSTRACT
The purpose of this study is to determine factors and identify the preferred type of financing among small business in rural areas in Tanzania. Written survey was used for data collection and the sample size was 102. Quantitative method was used to test the financing preference. It was found that business expansion (54%) and family obligations (32%) were the determining factors for external financing in the developmental stage. In the start-up stage of the business, the study finds that business owners preferred internal financing (61.7%) from their own savings for financing business. In the developmental stage, they preferred external borrowing from Savings and Credits Cooperatives Societies (SACCOS) (W = *161.8*), banks and other MFIs (*W=57.2)*. The small business owners' views SACCOS as most preferred type of financing due to its accessibility and impact on their businesses. The study concludes that internal and external funding is the backbone of rural MSMEs' financing in Tanzania. We recommend authorities to strengthen rural SACCOS to bring significant impact on rural business development.

Key word: Micro and small rural business, financing preferences, business cycle

CITATION:
Lubawa, Galinoma, Shirima, Andrew, and Nandonde, Felix Adamu (2018). "Financing Preference for MSMEs in Rural Tanzania." *Inter. J. Res. Methodol. Soc. Sci.*, Vol., 4, No. 1: pp. 18–35. (Jan. – Mar. 2018); ISSN: 2415-0371.

1.0 INTRODUCTION

In recent years Tanzania has witnessed the growing trend of Micro Small and Medium Enterprises (MSMEs) whereby the majority of them were established after the year 2000 (Liheta, 2016). Given their nature and characteristics of productivity, MSME is the main contributor to the economic development of many developing countries. For instance, it is estimated that MSMEs creates approximately 80% of new jobs in Canada and account for more than 63% of the GDP in developing countries (Ayyagari et al, 2005). In Tanzania, the sector employs over 33% of the labour force and it contributes to 27% of Gross Domestic Product (GDP) (Finscope Survey, 2012). There are about 3.16 million MSMEs in Tanzania whereby 54 percent of these operate in rural areas (Finscope Survey, 2012). It is estimated that 54% are owned by women and 46% by men (Finscope Survey, 2012). The role of MSMEs in Tanzania cannot be undermined as far as these enterprises create beneficial links between the non-farm economy and agricultural activities to alleviate rural poverty (World Bank, 2007).

There is a general consensus that access to finance and lack of collateral limit MSMEs in developing economies. To leverage resource challenges MSMEs in Tanzania prefer multiple loans (Finscope Survey, 2012; Mpogole *et al.*, 2012; Chalu and Lubawa, 2014; Lubawa & Louangrath, 2016). Despite of these efforts, however, the sector still faces challenges in accessing financial credit. These challenges include poor managerial skills, lack of skilled staff, uncontrolled growth, lack of government business support services, serial and historical traditions, high cost of inputs, advocacy capacity, inadequate quality control, and cultural environment (Cook & Nixson, 2000; Finscope Survey, 2012; Mbura, 2013). Rural areas are expected to face more challenges than urban business due to nature of poor business environment, risk credit market, poor infrastructure and the structure of rural economy. Kira and He (2012) observed that firms located in urban area in Tanzania have a higher possibility on access to loan or debt financing than firms located in rural area.

Despite contributions made by MSMEs toward economic development, the sector in Tanzania still faces the financing challenge (Mori et al, 2009, Finscope survey, 2012). This study investigated the most preferred type of financing and its reasons of funding MSMEs at start-up and development stage among business owners in rural Tanzania.

A sample of 102 small businesses was used. This paper examined the following issues: the start-up capital, age of the firm, owners' age and education level, reasons for external financing and the financing preferred sources in the growth period. The result from this study is expected to serve as a guideline to regulatory authorities on how to provide sustainable financing to small business in the rural regions.

The paper is divided into five sections. Section two reviews small business structure and financing in Tanzania, current sources of finance, and the finance issues and constraints faced by small businesses. Section three discusses theory relevant to small business financing. Sections four analyse the survey results and compare them to prior studies. The final section presents policy implications of the research.

2.0 MSME STRUCTURE AND FINANCING IN TANZANIA

The definition of MSMEs varies from country to country and from sector to sector. In Tanzania, the SMEs development policy of 2002 defines a micro, small enterprise as one with capital investment of between Tanzanian shillings (TZS) 5,000,000 and 200,000,000. According to the Small and Medium Enterprise Development Policy of the Government of Tanzania, the size categories of small businesses are defined through the number of employees and the capital investment in machinery (see Table 1).The number of employees includes the working business owner and manager. According to Tanzania Revenue Authority (TRA), SME is one whose annual taxable turnover is less than Tanzanian shillings 40,000,000. It should be noted that when an enterprise falls under more than one category (e.g., one employee but capital investment greater than TZS 5 million), the level of capital investment is the deciding factor in determining the size category. Schreiner and Leon (2001) pointed out that, a comprehensive definition of microenterprise should have three components namely; type of activity, investment limits, and number of employees. However, the definition of a 'small business' used in this study was taken to include all enterprises with capital investment of up to Tanzanian shillings 200,000,000. Based on SME policy, categories of SME definition in the country are classified as follows:

Table 1: Government of Tanzania's categorization of business size

Types of business	Employees	Capital Investment in Tanzania shillings
Micro business	1- 4	Up to 5 million

Small business	5 - 49	Above 5 million to 200 million
Medium business	50 - 99	Above 200 million to 800 million
Large business	100 +	Above 800 million

Source: URT Ministry of Industry and Trade (2003); 1 US$ is equivalent to Tanzanian shillings 2237.30 in April, 2017).

According to Finscope survey (2012), Tanzania has significantly more micro enterprises than small or medium sized businesses in all broad sectors (Table 2).

Table 2: MSME broad sector allocation by enterprise size

Size	Total	Trade	Services	Manufacture	Others
1 to 4 employees	3,074,736	1,710,884	942,596	406,426	14,830
5 or more employees	88,150	40,012	21,639	23,965	2,534

Source: Finscope survey (2012).

Finscope survey shows that (2012), there were about 2,754,697 people in Tanzania who owned and ran a micro, small or medium size business. While 86.2% of the small businesspersons owned and ran only one business, the rest owned and ran two (12.8%) or more (1%) businesses. In total, all MSME owned and ran about 3,162,886 MSMEs. About 96.4% of small business owners were sole proprietors, and about 3.5% were in partnerships in which 77% of all partners were related. About 1, 459, 167(53%) Small business owners are found in rural areas. About 80.6% of all small businesses had a business name, while only 3.9% of small businesses were registered under Business Registrations and Licensing Agency in Tanzania (BRELA). Although the majority of small businesses were not registered under BRELA many still held other licenses (e.g. Trade license, Professional registration, Local authority license, Daily license) that indicate some degree of formality, (see Table 6.6). A few small businesses had a tax identification number (TIN) for the business that may also serve as an indicator of formality and hence paid several taxes and levies, as shown in Table 3.

Table 3: Taxes and levies paid by MSMEs by area (%)

Item	Total	Dar Es Salaam	Other urban	Rural areas	Zanzibar
Income tax	5.2	7.7	8.2	2.8	14.6
Property tax	2.0	4.2	3.2	0.7	6.6
VAT	2.5	3.0	4.5	1.3	17.3
PAYE	1.4	3.1	1.3	2.3	2.5
NSSF/ZSSF	0.2	0.6	0.3	0.0	0.1
Excise levies	0.2	0.6	0.2	0.0	0.0
Other taxes/levies	2.2	4.1	3.6	1.0	8.8
Skills development levy	1.9	3.6	2.9	1.0	4.4
Unofficial" levies	0.9	2.0	1.4	0.3	3.3
Electricity/ water levies	3.0	6.0	6.0	5.2	9.5
Daily taxes	0.9	2.1	1.3	5.5	5.5
Business license	0.1	0.1	0.1	0.1	0.0

Source: Finscope survey (2012)

Finscope (2012) found that the motivation for starting a business for most MSMEs' owners originated from their need to support their families or to supplement their incomes. Less than a third of small business owners cited their achievement as a motivation for start-up. Two-thirds of small businesspersons were mostly motivated by the financial benefits they expected to generate. Before they started their business, many male small business owners had run a small business in another sector that gave them at least some small business experience. However, most of them were hampered by capital constraints.

Finscope (2012) also indicated that the highest level of education achieved by MSME owners is primary school, with male business owners generally obtained higher levels of education than female owners; this was particularly pronounced in rural areas. However, education has been partly improved through attending short training courses after their formal education. Consequently, most small businesspersons regarded themselves as either unskilled or semi-skilled (Finscope survey, 2012). MSMEs' owners were found in all age groups, with the greatest concentration in the 25-34 year old group (Table 4).

Table 4: Ownership by age groups and gender (%)

Age group	Total	Male	Female
Under 25 years	11.0	10.4	11.6
25 to 34 years	36.0	38.4	34.2
35 to 44 years	30.7	29.4	31.8
45 to 54 years	15.5	15.4	16.0
55 years and above	6.8	6.8	6.4

Source: Finscope survey (2012)

Small businesses are considered as important foundation for the growth of private sector in Tanzania in term of job creation, income generation, feeding inputs to large-scale industries, development of domestic technology, utilization of local resources and stimulation of growth in both urban and rural areas (Liheta, 2016). Through business linkages, partnership and sub-contracting relationships, small businesses have a great potential to complement large industries. For a developing country, a strong and productive industrial structure can only be achieved where small and large enterprises co-exist and work in partnership.

In Tanzania, small business enterprises are found in food processing, textiles, edible oil production, woodworks enterprises, retail shops, etc. that requires simple technology and local raw materials are abundant. These enterprises help to encourage the use of local resources for its production and thus can create a self-reliant economy. Small firms are also serving as ancillaries to large firms by providing parts, components, and accessories (Liheta, 2016). According to Finscope Survey (2012), the majority of small business owners started their business to support themselves financially and few of them were driven by entrepreneurial motives.

An adequate financing for small business is important in every stages of business cycle. However, in practice, the microfinance interventions are accessible only to the small businesses already existing in the market and ignores the new entrants in the market particularly those in the rural areas who are looking for start-up capital. They forget that no business entity, whether large or small, can be incorporated, established and expanded without adequate start-up capital. For most small businesses in Tanzania, start-up capitals are funds provided by the start-up team, family and friends, prior to and at the time of the inception (Mpogole et al, 2012; Chalu and Lubawa, 2014).

The literature confirmed that small business owners' tend to rely heavily on their past savings, followed by informal sources of credit from family or friends, particularly at business start-up stage (Aryeetey et al., 1994; Paul et al., 2007; Osei-Assibey et al, 2012). the longer firms stay in business operation, the more it signals that it can weather tough economic conditions Chandler (2009), and that the use of external debt financing particularly bank loan becomes more common and most preferred once the business is up and running (Wyer et al., 2007). Berger and Udell's (1998) financial growth model, shows that small business have a financial growth cycle in which financial needs and options change as the firm starts-up, grows, gains further experience and becomes more transparent with its information. As firms continue to expand, they could access to public equity and debt markets. However, in order to start or grow any form of business, financing is essential. There are various sources of financing available to small firms, depending on a variety of factors. In Tanzania, small business can raise funds through formal and informal financial institutions. Through a review of previous studies on MSMEs financing in Tanzania, we found that most of studies identified the following financing sources for small business in Tanzania;

2.1 Internal sources of financing

Internal financing or self-financing comes from individual saving it is constrained by the low savings capability arising from low incomes. Usually, most savings are held in no monetary forms, such as livestock, grain, maize and other such trees and land, though it could be also as the accumulation of cash. It should be noted that savings in rural Africa is often not clearly distinguishable from investment or consumption, particularly where it involves animals or consumables (Liheta, 2016). It can be argued that savings is a source of equity financing (self-financing) to most of small businesses. Frederick et al (2006) and Cassar (2004) explained that equity financing is the best financing option for starting a business. It should be noted that, small businesses tend not to go for equity financing from the public in order to keep the original of ownership (Fredrick et al, 2006), and due to cost burden (admission fees, advisors and broker commissions), lack of education about the process of listing (Nassr and Wehinger, 2015).

2.2 External sources of financing

As the business grows, the trade-off theory states that a firm can derived optimal debt ratio through a trade-off between cost of bankruptcy and tax advantage of borrowing (Scott, 1997). The higher the profits, the lower the expected costs of financing distress and hence the business would prefer debt financing over equity. Although the theory is commonly applied to larger firms; however, its usefulness is also important in small firms. Myers (1984) and Fazzari et al (1988) argued that small business will choose external funds than internal to normalize its business investment if larger cash flows are needed.

In Tanzania, microfinance scheme is a major external funding source for most MSMEs. The scheme provides microcredit to small businesses that are not acceptable by the commercial banks due to lack of collateral. The microfinance scheme tends to use the credit methodologies that simply employ collateral substitutes and innovative recovery practices. The credits methodologies provides terms and conditions that are flexible and easy to understand and are suited to MSMEs (Liheta, 2016).

2.3 Microfinance Institutions (MFIs) in Tanzania

Tanzania has many Microfinance Institutions (MFIs) that give loans to small businesses. The majority of these MFIs were started when the government developed the National Microfinance Policy in line with financial sector reforms in 1991. The National Microfinance Policy was launched on February 2001 with key objective of establishing a basis for evolution of an efficient and effective micro finance system that served the low-income segment of the society. The policy also spelt out the national vision for the development of Micro finance as a tool for poverty eradication through the widespread access to micro financial services across the country, on a commercially sustainable basis, thereby contributing to the economic growth of Tanzania. Since then MFIs in Tanzania has been providing financial services to the unbankable community in the rural and urban areas. Some of the services offered by MFIs are; macro savings, micro credits, micro insurance and money transfer. MFIs in Tanzania have created several impacts to beneficiaries. Some scholars report positive effects of MFIs; for instance; some studies revealed the increase of clients' expenditure, accumulation of assets, improvement of health, education, food security, nutrition and business growth for the poor(Kessy and Urio, 2006; Kessy and Temu,2009; Stewart et al,2010). Other studies report adversely, such as confiscation of borrowers' assets which might lead to poverty, and sometimes stress and mental burden, increase of workloads for women (Nghiem, 2009; Mpogole et al, 2012; Kato and Kratzer, 2013). In general, MFIs have been the major success for the business to compete and for poverty alleviation (Christopher, 2008). Although their reach to broader population and small scale businesses is still limited due to concentration of MFIs in selected regions.

The followings are MFIs operating in Tanzania apart from SACCOS which dominant group among MFIs, other MFIs dominant in issuing loans to small businesses are; Foundation for International Community Assistance (FINCA),Small Industries Development Organization (SIDO), Promotion of Rural Initiatives and Development Enterprises (PRIDE - Tanzania), Small Enterprise Development Agency (SEDA), YETU Microfinance, financial Non-Governmental Organizations (NGO), and Savings and Credit Associations (SACAs). Other MFIs are; Bangladesh Rural Advancement Committee (BRAC), Microenterprise Development Agency (MEDA) and BAYPORT.

There are also commercial banks which have down-scaling of their commercial services offered so as to allow provision of microcredit similar to what is currently offered by MFIs. These banks are: The National Microfinance Bank (NMB Plc), AKIBA Commercial Bank, CRDB Plc, Tanzania Postal Bank and Equity Bank Tanzania. The NMB encourages and expands microfinance in three ways; (i) Loans to micro and small enterprises for the purchase and inventory and supply of goods, (ii) Collection and payment services to large corporate clients to/from micro and small enterprises and (iii) add-on services such as money transfers and payroll services to both the large corporate clients and micro and small enterprises. AKIBA currently offers several microfinance loan products tailored to suit every type of customer. The microfinance products offered at AKIBA are savings deposits, group (called Solidarity Group) and individual microenterprise loans (called Biashara Loans). CRDB is a more conservative bank reduces its risks in microfinance by loaning to groups such as SACCOs instead of micro and small enterprise borrowers. CRDB provides SACCOs with financial stability.

The Tanzania Postal Bank (TPB) is also the commercial bank that is involved in microfinance. The TPB was created by an Act of Parliament and is under the Companies Registrar and under Tanzania Postal Bank Act No.11 of 1991 as amended by Act No.11 of 1992. The bank provides microfinance services to Small and Medium Enterprises (SMEs). These are small and medium size loans available to Small and Medium Enterprises to support their financial requirements including enhancement of working capital and financing investment related activities. The objective is to expand business operations and grow their profitability to become viable and sustainable businesses. The facility may be issued in form of an installment loan or an overdraft facility. Equity Bank Tanzania also provides Micro Business Loans (e.g.group loan that is a credit product specially designed for group borrowers or customers with micro businesses) and SME-Business Loan (available to SME's operating in Transport, Trade and Commerce, Construction, Manufacturing, Education, Health and other service sectors). However, most banks and other financial institutions in Tanzania are located in urban areas (Facet, 2011) that that results in difficulty in accessing microfinance service by small scale businesses in rural areas (Wangwe & Lwakatare, 2004).

In rural Tanzania, there are semi-formal microfinance institutions and Non-governmental organizations providing microfinance services (Nyamsogoro, 2010; Facet, 2011). Savings and Credit Cooperative Organizations (SACCOs), RoSCAs, SACAS, the Village Savings and Loan Associations (VSLAs), Village Community Banks (VICOBA) and community conservation banks (COCOBA) are some examples of semi-formal microfinance institutions. The semi-formal microfinance institutions in rural areas are made up of farmers and non-farm members offered micro credit provision to members only. The credit can also be granted in financing basic needs like education, agriculture and health. They provide savings and internal lending for communities (Facet, 2011; Randhawa & Gallardo, 2003; Rubambey, 2005). The study by Flora (2015) shows that small scale businesses in rural Tanzania have access and benefited from microfinance services from different MFIs.

SACCOS are co-operatives microfinance institutions using co-operative values and principles in provision of financial services. SACCOS are dominant in microfinance market. They are owned, governed and managed by its members. They usually require membership fee, and/or share capital or obligatory savings from all members (Liheta, 2016). In Tanzania, they are regulated and supervised by the co-operative act. Since became in operations, rural SACCOS in Tanzania have great assisted in addressing the financing problem for rural borrowers (Wangwe, 2004; Kessy and Urio, 2006; Kessy and Temu, 2009; Qin and Ndiege 2013). The Ministry of Finance of Tanzania reported that until March, 2013 in country the number of registered SACCOS was about 5346 while amount loaned reached shillings 627.2 billion in 2011(MOFT, 2012; 2013). Since their formation started in the 1980's, after the adoption of the free market liberalization policies, SACCOS has been very important microfinance institution to community. Literature evident that SACCOS are very helpful MFIs in serving rural Tanzania where most of people are unbankable Wangwe (2004). A Study by Bwana and Mwakujonga (2013); Quin and Ndiege (2013) found that SACCOS contribute to about 40% of the Tanzania's GDP and play important roles for financing small and medium enterprises in the rural areas. Magali (2013) also realized that rural SACCOS improved the livelihood of members on education, health, physical assets, crop yields and business capital. SACCOS play vital role in boosting the social-economic development of rural areas.

2.4 Sources of Financing by Life Cycle Stages

Like small firms around the world, micro and small business in Tanzania also undergo several stages from inception to maturity. This life cycle begins in the developmental stage through various stages of revenue growth. However, these stages depend on economic environment surrounding the firms. Berger and Udell (1998) asserted that the particular phase of business's life cycle determines the nature of its financial needs, the availability of financial resources and the related cost of capital. Micro and small businesses are thought to have a financial growth cycle in which financial needs and options change as the business grows, gains further experience, and become less informational opaque. Table 2 shows life cycle stages, types of financing for each stage, sources of financing and entrepreneurial process activities.

Table 2: Sources of Financing by Life Cycle Stages

Life cycle stage	**Types of financing**	**Major sources/players**	**Entrepreneurial process activities**
Development	Seed financing	Entrepreneur's assets Family and friends Bootstrapping	Developing business idea Developing opportunities
Start-up	Internal financing	Internal funds Entrepreneur's assets Bootstrapping	Gathering resources Building operations
Expansion	External financing	Business operations Suppliers and customers Obtaining MFIs loans	Managing and building operations
Maturity	External financing	Business operations Suppliers and customers Obtaining MFIs loans	Managing and building operations

During the developmental stage of a business's life cycle, the primary source of funds is in the form of seed capital to determine whether the idea can be converted into a viable business opportunity. The primary source of funds at the developmental stage is the business owner's own assets and sometimes financial bootstrapping; that is, creative methods, including barter, to minimize the cash needed to fund the business (Leach and Melicher, 2010). In the business start-up stage, it is a common practice for micro and small business owners to use internal funding capital in order to minimize financing cost (Berger and Udell, 1998).

2.5 Constraints Faced By Small Business in Tanzania

Most commercial banks consider SME's as risky borrowers. Nyamsogoro(2011) on his study on "factors influencing financial sustainability of rural microfinance institutions in Tanzania", concluded that; microfinance capital structure, interest rates, difference in lending types, loan repayment plan, cost per borrower, product type, microfinance size, and the number of borrowers are significant factors affecting rural microfinance institutions in Tanzania. Based on these facts, one may assume that, providing financial services to the rural area has always remained a challenge to MFIs due to the inherent difficulties associated with low population density, seasonality, isolated markets, lack of adequate physical and human infrastructures (e.g. roads, electricity, poor telecommunication services and limited technology, education and health), that's why most of MFIs in rural are small, of which are member-based cooperatives, (e.g. SACCOSs and SACAs). Finscope-Tanzania survey (2012) reported that majority of adult population in rural areas do not use banks products and only 18.4% of the population are covered by formal financial institutions. Kira and He (2012) pointed out that among other factors associated with firms' characteristics that have impacts on small business is access to external financing from different sourcing of financing in Tanzania are, age of firm, and location. According to Kira and He (2012) firms located in urban areas in have a higher possibility on access to debt financing than firms located in the rural area and that older firms find easier to access debt financing than younger firms. The study also found that firms located in urban areas have higher possibility on access to loan or debt financing than firms located in the rural area. It has also been suggested that participation in entrepreneurial activities is influenced by level of education (Peter, 2001), and that people with a low level of education have more difficulties finding a paid job and; therefore, see no other possibility than to be self-employed in entrepreneurship (Wit & Van 1989). Olomi (2005) pointed out that those who unable to find jobs in the formal sector end up in the informal sector whereby MSMEs are dominating. This is a crucial issue to be concerned particularly in Tanzania where about 80 percent of the population are in rural location and are excluded from banking services (Rubambey, 2005).

Most entrepreneurs in rural areas are faced with financial illiteracy, awkward geographical location (require special journey) and temporal constraints in accessing formal financial services. Berger and Udell (2002) pointed out that the geographical closeness between lenders and customers has significant association for customers to access loan. The MFIs which are geographically close to MSMEs are capable of utilizing soft available qualitative information for credit analysis and evaluation. To address these challenges, saving-up strategies come in especially where the other mechanisms are not available. Savings are accumulated amount of money in a safe place at home (home box) until they have grown into a useful large sum (Rutherford, 1999). Savings as a micro-finance factor enable people with few assets to save, since they could make weekly savings, as well as contribute to group savings, and such savings are mobilized by the micro-finance institutions for further lending to other clients (Mkpado & Arene, 2007). Household saving at home is arguably the most relevant savings mechanism in Kilolo villages since it has no entry barriers, and has no transaction cost, and impose no interest charges (Kilolo District Council, Social Economic Profile, 2013). In a tribal language, this saving strategy is known as "*fiyite amapesa*" (that means savings the money in a secret place at home for future or emergency use). Using private savings does not bring additional obligations and the decision to use them can be made without reference to a third party. With savings approach, the terms of contribution range from a few days or weeks to many years depending on the entrepreneur's objectives mostly for business motivation. Traditionally, rural areas tend to build lump sums seasonality after the sale of harvest or animals (Kilolo District Council, Social Economic Profile, 2013). Other form of savings which is also prevalent in many rural areas of Tanzania is savings in-kind. This is when an individual opts to use physical assets as a store of value and use it in future when there is a need of liquid cash. This may take form of animals, maize, beans, and forest product. However, the saving method, whatever form it takes, is faced with higher risks like theft, produces are susceptible to disease, destructive tendency of moisture and damage by rodents and consumption to meet daily life cycle needs and unanticipated financing emergencies (such as illness, accident, funerals and fire). Thus, the use of personal savings for individual growth as well as contribute to group savings, and such savings are mobilised by MFIs for future borrowing to is linked to the Pecking Order Theory(POT). According to POT, (see Abor (2008) and Hussain and Matlay (2007)), the order of the preference is from the one that is least sensitive (and least risky) to the one that is most sensitive (and most risky) that arise because of asymmetric information between corporate insiders and less well informed market participants. This is because internal financing incurs neither security issuing or flotation costs nor disclosure of financial information and, thus no transaction cost. According to the POT, if the external financing is needed, then debt, which is associated with less severe information asymmetry, is preferred over equity issue. Financing decisions of MSMEs are better explained by the POT when faced with information asymmetry.

3.0 RELEVANCE OF PECKING ORDER THEORY FOR SMALL BUSINESS FINANCING

POT was developed for large public companies; however, it is equally appropriate to small firms because the 'Pecking order hypothesis' in keeping with the prior that external borrowing is by far the largest source of external finance for small businesses (Scherr et al,1990; Cassar and Holmes,2003). The issuance of new equity for small business is questionable due to nature of firm's characteristics. Most of these firms are family businesses, that small business' manager tend to be the business owners and they are not interested to dilute their ownership claim (Holmes and Kent, 1991, Jordan et al, 1998). This implies that external equity financing for small business is inappropriate. Ibbotson et al, (2001) argues that POT is more relevant for the small business sector because of the relatively greater information asymmetries and the higher cost of external equity for small business. The information asymmetry and agency conflicts are likely to arise in small business simply because of their characteristics, i.e. being controlled by individual or family or related people. Scherr et al, 1993 also noted that the costs of asymmetry are more evident for small business than for large publicly traded companies, making differences in costs between internal equity, debt, and external equity consequently greater. Therefore, POT financing preferences may be more relevant to small firms than large ones. The literature suggests that small business owners source their capital from a pecking order: first, their own money (retained earnings); second, short-term borrowings, as liquidity management of small business is crucial if the survival and prosperity of small business is to be ensured (Deakins et al, 2000; Sardakis et al, 2007); and third, long-term borrowing. The Pecking Order Theory developed by Myers and Majluf (1984) and Myers (1984) does not predict an optimal capital structure.

Against this background, rural MSMEs financing remains less studied area in Tanzania. Empirical evidence supports the applicability of the POT in explaining the financing of MSME which has been done in foreign countries where business environmental might be different from rural environment in Tanzania. Few studies in developing countries showed contradicting results. While some studies (e.g. Green et al, 2002; Gebru, 2009; Osei-Assibey et al, 2012) argue that MSME's financing preference generally conforms to POT, others (e.g. Murray and Goyal, 2003) have shown that POT fails where it should hold, especially for small firms where information asymmetry is a problem. Thus, little is known about rural context in Tanzania. The main objective of the study was to provide some preliminary descriptive evidence of rural MSME financing. This study, therefore, took the MSMEs in the wards of Ilula, Mtitu and Ruaha Mbuyuni in Kilolo District as a case study to investigate the rural MSMEs Financing Preferences: does POT hold? To achieve this end, a questionnaire survey was administered to a sample of 102 SMEs owners' in the target areas by examining issues including the start-up capital, age of the firm, owners' age and education level, reasons for external financing and the financing pecking order in the growth period.

4.0 STUDY LOCATION AND METHODOLOGY

The study was conducted in rural Tanzania in the District of Kilolo. The location was purposively selected for two major reasons. First, the presence of active microfinance institutions (e.g. SACCOS) likes other SACCOS in the country, where microfinance services have been growing rapidly (MBF, 2010). There are eleven (11) registered active SACCOS in Kilolo rural, whereby; only seven (7) SACCOS are operating have shown to be more active. A total of Tshs. 790,289,154 was deposited in eleven (11) SACCOs accounts by the year 2012. At the same time, a total of Tshs. 4,911,509,650 was credited to individual members in 2012. Second, despite the fact that there is a tremendous deposit in SACCOS, however, by the end of 2012, NMB, Tanzania Postal Bank and FINCA were the only financial institutions that were operating in the district (Kilolo District Council, Social Economic Profile, 2013), however, savings practices are still popular in the District.

In this study, data was collected in April / May 2016 by using a survey questionnaire which targeted small businesses. A sample of 102 entrepreneurs from three wards was included in the survey. All questions in the survey were close-ended to limit the respondent's answers to the survey questions and encourage their focus on the issue at hand. Descriptive statistical methods and employed to analyse and explain the data. Sample size was estimated by using the following formula (Amin, 2002).

$$n = \frac{Z^2 \sigma^2}{E^2} \quad (1)$$

where n = sample size; σ = estimated population standard deviation; and E = standard error determine by $E = \sigma / \sqrt{n}$ assuming confidence interval of 95% for the estimated population proportion, maximum error of 5%. A test sample of 30 was used to calculate minimum sample size according to equation (1). The result showed a minimum sample of 27 was required. A sample of 102 counts was collected for this study.

5.0 RESULTS AND DISCUSSION

5.1 Description of respondents

Respondents involved in this study were business owners from the selected areas. The results indicate that most of the respondents came from age of between 18 to 39 years (76%). The age is an indicative that people can have extra years of saving and accumulating a business capital. The findings are consistent to other studies by Tundui (2012); Chalu and Lubawa (2014), which noted that most of the small business owners in Tanzania were younger than 40 years. Other respondents were in the age group of 40-49(18%), and 50 years and above group was the smallest (6%). These were people born before 1980, this is the time where Tanzania followed socialism. The country embraced a state-controlled economy. There was no room for market economy due to dominance of Marxist-Lenninist oriented social, political and most importantly economic thinking. During this period, people experienced an anti-entrepreneurial educational, political, legal policy and regulatory environment where an attempting to create free private enterprises as more or less equivalent to being an enemy of the state. An indication that either the elderly people freely chose not to engage themselves in the small business due to poor entrepreneurial skills or thus debt market discouraged them to avoid risk of input repayments failure. This finding indicates that individual attributes (such as age) may play an important role in entrepreneurial activity. Education was found to be among the factors explaining the owners' financing preference. Generally, the results revealed that most of MSMEs' owners in rural areas possesses relatively low education, 50 percent of them had not completed secondary education. As expected, findings revealed that business owners with lower education level (60 percent) are interested to abide with explanations of Pecking Order Hypothesis, which exhausts all internal financing sources before opting for external financing. Educated MSMEs owners' are in good position to link between the risk and cost associated with source of financing mainly debt with intention of optimizing the value of the firm and which decides for an optimal capital structure. The results also showed that 46% of respondents owned business with capital of less than Tanzanian shillings 5,000,000.

In case of the age of the business, the results indicate that 46 firms (about 45.1%) have been operating for more than ten years. This is an indication that most of these firms can go on with operations due to experience on encountering challenges especially financial challenges. As more time passes and enterprise matures, reputation is built and confidence is gained, the desire for long-term growth and expansion becomes inevitable. Chandler (2009) pointed out that; the longer the firm stays in operation, the more persistence to unpleasant economic circumstances. It was expected that external finance would be preferred or desired from debt market, since the firms then become more acceptable to the markets. Firms with less than 5 years in operation are less likely to rely on debt financing from debt market (Klapper, 2010).

Table 3.1: Description of Respondents

Age(years)	**Frequency**	**Percent (n = 102)**	**Education**	**Freq.**	**Percent (n = 102)**
18-28	18	17.6	No formal education	27	26.5
29-39	56	55.0	Primary education	41	40.2
40-48	20	19.6	Secondary education	16	15.7
49 and above	08	7.8	Certificate or diploma	14	13.7
			Degree and above	4	03.9
Total	**102**	**100**	**Total**	**102**	**100**

Years in business	**Freq.**	**Percent (n = 102)**	**Size by Capital (TZS)**	**Freq.**	**Percent (n = 102)**
Less than one year	05	05.0	Below- 4,999,999	52	51.0
1 – 5 years	23	22.5	5,000,000 - 9,999,999	19	18.6
6 – 10 years	28	27.4	10,000,000 - 14,999,999	16	15.7
More than 10 years	46	45.1	15,000,000 - 19,999,999	11	10.8
			20,000,000 and above	04	3.9
Total	**102**	**100**	**Total**	**102**	**100**

Note: Percent = Frequency / 102

Respondents were also asked about the financial sources at start-up (see table 3.2). About 61.7 percent of them indicate that the key financial sources at the business start-up come from private savings, 21.6 percent borrowed from friends and relatives, 16.7 percent borrowed from MFIs and 0 percent from commercial bank and government support. This is consistent with other studies in foreign countries. For example, Bank of England (2003) data suggests that 60 percent of business owners used their personal financial resources and finance from family members as the main sources of capital to start their business. Berger and Udell (1998) pointed out that, at start-up, small firms tend to rely on initial finance from owner's savings, family and friends.

The findings confirmed that rural areas still are excluded from banking services due to strict requirements for accessing the funds from a particular financial institution: Formal business address, having a good, convincing and bankable business plan, having collateral, showing good business records, including sales, revenue and profit, indicating reputable referees, having to for loan processing fees, paying high interest rates and at times rough treatment in case of delayed payment or default. All these and many others factors are hindrances for the rural MSMEs to access finance. They are understandable from a business perspective, but they remain a necessary evil for small business. The findings also show that the Government support in rural area is still limited. Thus, private savings remains the only source of funds to most rural entrepreneurs due to constraints pose by the financial institutions in one hand and the rural accessibility in other hand.

We asked respondents to mention one major reason for external financing. Results show that business expansion (54%) was the major reason for rural MSMEs external financing. This is an indication that rural firms attempts to grow despite of all geographical and financial barriers. Family obligations (32%) ranked second after that of business expansion. This means that although they intended to take loans for business running, intentionally used for family obligations. This is an indication that rural entrepreneurs did not distinguish business from family matters due to low income. Olomi (2009), urged that the lack of distinction between family obligations and business itself in Tanzania has been persisted for quite some time. Other reasons reported were financing deficit (8%) and developed a new business (6%). we can conclude that innovation is still unthinkable to rural small businesses.

Table 3.2: Sources of Start-Up Capital and Reasons for External Financing by SME's Owners

Start-Up Capital			**Reasons For External Financing**		
Category	**Freq.**	**Percent (n = 102)**	**Category**	**Freq.**	**Percent (n = 102)**
Commercial bank loan	0.0	0.0	Business expansion	52	54
SACCOS /MFIs loan	17	16.7	Financing deficit	11	08
Government support	0.0	0.0	Developed a new business	06	06
Friends and relatives	22	21.6	Family obligations	33	32
Own Saving	63	61.7			
Total	**102**	**100**	**Total**	**102**	**100**

Note: Percent = Frequency / 102

5.2 Preference of financing in Growth Period

To test the Pecking Order Hypothesis during the growth period, a question was designed to solicit rural MSME's owners about their financial preference during the growth period of their firms. They were free to list their three financing preference sources as first choice, three sources second choice and three sources as third choice, by mentioning the names of financial source whether commercial banks, SACCOS/MFIs or and equity.

The choice of financing preferences entails an important decision for every firm whether public or private, large or small. In fact, considering its duration, amount and irreversible character, a financing decision is regarded as a major and strategic one. Therefore, the process of a financing decision should be conveniently modeled. For this growing stage, the multiple weighted scoring model (MWSM) was adopted. The model is the best known and simplest Multi-Criteria Decision Analysis (MCDA) method for evaluating a number of alternatives in terms of a number of decision criteria by screening, prioritising or selecting the alternatives based on human judgment from among a finite set of alternatives in terms of multiple usually conflicting criteria (Hwang and Yoon, 1981). We assume that all alternatives score with respect to all criteria are known or has been estimated by the decision maker.

According to multiple weighted scoring method, each financing preference is ranked by its relative importance (mapping scale) (Table 3.3).

Table 3.3: Financing Preference and Weighted Importance

Financing Preference	Importance weight (b_i)
1st Preference	3
2nd Preference	2
3rd Preference	1

Score for all choices (frequencies (f_i)) would be calculated using equation (2).

$$\sum_{i=1}^{n} f_i = f_1 + f_2 + \ldots + f_n \quad (2)$$

where f_i = score in terms of preferences and or type of financing source i = 1, 2,…,102, and n = sample size (102).

Individual weighting values (W_i) for financing preference were developed according to equation (3). Multiply importance weights (b_i) by scores to arrive at a weighted score for each criterion (W_i). The computed sets of numerical weights determined the relative importance of the evaluation categories. Then the weighted scores are used to arrive at an overall project score: the final score for each project becomes the sum of all its weighted criteria. The highest weighted score ranked (R) as the most preferred financing by respondents.

$$W_i = \frac{\sum_{i=1}^{n} f_i b_i}{n} \quad (3)$$

where, W_i = weighted score for individual preference, i = 1,2,3,..., k, k = number of score for each financing preference, and n = sample size (102). The weighted score vector must satisfy the condition that:

$$\sum_{i=1}^{3} W_i = W_1 + W_2 + W_3 = 1 \quad (4)$$

where, W_i = weight of criterion financing preference either first, second or third choice and $W_i \geq 0, i(1,2,3)$.

Table 3.4 points out that the first financing preference by rural MSMEs follows the pattern SACCOS (161.8), Banks (57.2), Internal financing (56.7) and equity (0.0). Both SACCOS and banks are external financing. This proves that on growing stage the rural MSMEs financing strategies would prefer external debt over internal. The findings reaffirm POT to establish a hierarchy of financing preference.

The MSMEs owner's criteria for preferring SACCOS financing was that loans by SACCOS can lead to significance impacts on business. Its smoothing accessibility was another criterion. The accessibility was fair because the borrowing comes from client's contribution. The findings stressed the evident of Wangwe (2004) that SACCOS are important MFIs in rural Tanzania where most people are not served by formal financial institutions.

Table 3.4: Financing Preferences in growth period

Preferred Financing Sources									
		External Source				**Internal Source**			
Financing source		**SACCOS**		**Banks/ MFIs**		**Internal Financing**		**Equity**	
	$\mathbf{b_i}$	$\mathbf{W_i}$	$\mathbf{f_i}$	$\mathbf{W_i}$	$\mathbf{f_i}$	$\mathbf{W_i}$	$\mathbf{f_i}$	$\mathbf{W_i}$	$\mathbf{f_i}$
	3	69.6%	71	17.6%	18	13.7%	14	0.0%	0
	2	25.5%	26	21.6%	22	23.5%	24	0.0%	0
	1	4.9%	5	60.8%	62	62.7%	64	0.0%	0
$\sum \mathbf{f_i}$			102		102		102		0
$W_i = \frac{\sum_{i=1}^{k} f_i b_i}{n}$		**161.8**		**57.2**		**56.7**		**0.0**	
Creteria($\mathbf{C_i}$)		Accessibility, Positive impact on business		Difficult in accessibility, collateral demanded		Low package, low risk		Long period of waiting the dividend, Require huge capital for better returns	
Ranking		**1**		**2**		**3**		**4**	

Similar findings were reported in the study by Gebru (2009) and Osei-Assibey et al, (2012) in rural financing in Ethiopia and Ghana respectively. Both studies reaffirmed the POT to a hierarchy of financing preference. Furthermore, factors such as owner's age and level of education, age of business, size by capital, reasons for start-up capital were found to be major determinants of rural SME owners' financing preferences. In the presence study, we also found that these factors are also determinative for POT.

6.0 CONCLUSION AND RECOMMENDATION

In this paper, a number of issues surrounding rural MSME owners' financing preference in light of the Pecking Order Theory (POT) were investigated. The overall findings provide a significant support for the Pecking Order Theory. Factors such as start-up capital, age of the firm, owners' age and education level, reasons for external financing and the financing pecking order in the growth period are found to be major determinants of rural MSMEs owners' financing preferences.

We observed significant implications for the government and other stakeholders on rural development. The findings urge government and other policy makers to develop appropriate policies to resolve the problems of finance in the rural MSME sector. This requires the government and the banking sector to develop solid strategies to support rural MSMEs in financing. Moreover, the results also alert policy makers to re-examine the current credit guarantee institutional scheme in Tanzania, particularly SACCOS and give flexible attention to rural MSMEs. Possibly, rural MSMEs owners' could benefit from optimal capital structure education and financial management training.

This study was limited by the relatively small sample, nevertheless the research outcomes, , contributed to expanding the body of knowledge about rural MSME financing in Tanzania and could encourage further research on the role of financial management education to rural MSMEs owners', since POT was developed for large publicly companies and believed to be equally appropriate to small firms. However, the findings can also be used to provide basic information and stimulate discussion to on the need for a modified POT that could more reflect the special context of rural MSMEs financing in developing countries. We need to create an integrated rural financial system that is more responsive to varied financing needs of MSMEs at all stages. For developing economies, country's wealth depends on the vibrancy and dynamism of its MSMEs.

REFERENCES

Abor, J. (2008). "Determinants of the capital structure of Ghanaian firms", Research Papers RP_176, *African Economic Research Consortium*, Nairobi.

Ayyagari,M., T. Beck, and A. Demirgüç -Kunt, "Small and medium enterprises across the globe," World Bank Policy Paper, 2005.

Aryeetey, E. and Udry, C. (1997). "The characteristics of informal financial markets in sub-Saharan Africa", *Journal of African Economies*, Vol. 6, pp. 161-203.

Bank of England (2003). *Quarterly Report on Small Business Statistics*, Bank of England, October, available at: www.bankofengland.co.uk/

Berger, A., & Udell, G. (2006). A More Conceptual Framework for SME Financing. *Jouranl of Banking and Finance*, **30(11**), 2945-2966. http://dx.doi.org/10.1016/j.jbankfin.2006.05.008.

Berger, A.N. and Udell, G.F. (1998). ''The economics of small business finance: The role of private equity and debt markets in the finance growth cycle'', *Journal of Banking and Finance*, Vol. 22, No. 6-8, pp. 613-73.

Berggren, C., Olofsson, C. and Silver, L. (2000). ''Control Aversion and the Search for External Financing in Swedish SMEs'', *Small Business Economics,* Vol. 15 No. 3, pp. 233-42.

Bwana, K. M., & Mwakujonga, J. (2013). Issues in SACCOS Development in Kenya and Tanzania: The Historical and Development Perspectives. *Developing Country Studies*, 3**(5),** 114-121.

Cassar, G. (2004). The Financing of Business Start-Ups. Journal of Business Venturing, Vol. 19, pp. 261-283.

Cassar, G. and Holmes, S. (2003). ''Capital structure and financing of SMEs: Australian Evidence'', *Accounting and Finance*, Vol. 43, pp. 123-47.

Chandler, J. G. (2009). Marketing tactics of Selected Small Firms in the East London CBD Area. South Africa: University of South Africa.

Chalu, H. & Lubawa, G. (2014). An Empirical Analysis of SMES' Owners Perceptions on the Effects of Multiple Borrowings on Organizational Performance. *Paper Presented to 14th International Conference African Entrepreneurship & Small Business (ICAESB 2014);* White Sands Hotel May 29th and 30th, 2014, Dar es Salaam, Tanzania.

Christopher, Idowu Friday. (2008). Impact of Microfinance on Small and Medium-Sized Enterprises in Nigeria: Wuhan, PR China: School of Management, Wuhan University of Technology, 1864-1871.

Cook, P., and Nixson, F. (2000). Finance and small and medium-sized enterprise development. *Journal of Developmental Entrepreneurship,* **6**(1).

Facet, Triodos. (2011). Tanzania Microfinance Country Scan, Final report. Zeist, The Netherlands, Washington.

Finscope Survey (2012). "National Baseline Survey Report for Micro, Small, and Medium Enterprises in Tanzania". Ministry of Industry and Trade, Dar es Salaam; Financial Sector Deepening Trust (FSDT).

Frederick, H.H., Kuratko, D.F. and Hodgetts, R.M. (2006). Entrepreneurship: Theory, Process and Practice. Asian-Pacific ed. Australia. Cengage Learning.

Flora, K.A. (2015). The Challenges Confronting Small Scale Businesses in accessing Microfinance Services from MFIs Case Study: Rural Tanzania. *International Journal of Academic Research in Business and Social Sciences*, Nov 2015, Vol. 5, No. 11, ISSN: 2222-6990.

Galinoma Lubawa & Paul Louangrath (2016). "Using Altman Z-Score to Assess the Financial Effects of Multiple Loans on SMEs." *Inter. J. Res. Methodol. Soc. Sci., vol.* 2, no. 1 (Jan - Mar. 2016); pp. 63-86. ISSN 2415-0371.

Gebru, G.H. (2009). "Financing preferences of micro and small enterprise owners in Tigray: does POH hold?" *Journal of Small Business Enterprise Development*, Vol. 16 No. 2, pp. 322-34.

Green, J.C., Kimuyu, P., Manos, R. and Murinde, V. (2002). "How Do Small firms in Developing Countries Raise Capital? Evidence From A Large-Scale Survey of Kenyan Micro And Small Scale Enterprises", Economic Research Paper No. 02/6 Centre for International, Financial and Economics Research, Loughborough University, Loughborough.

Holmes, S. and Kent, P. (1991). "An Empirical Analysis of the Financial Structure of Small and Large Australian Manufacturing Enterprises." *Journal of Small Business Finance,* **1 (2),** pp. 141-154.

Hussain, J. and Matlay, H. (2007). "Financing Preferences of Ethnic Minority Owner/Managers in UK." *Journal of Small Business and Enterprise Development,* Vol. 13 No. 4, pp. 584-99.

Hwang C. L., and Yoon K. (1981). *Multiple Attribute Decision-making: Methods and Applications*, Springer, Berlin.

Ibbotson, R. G., Sindelar, J. L. & Ritter, J. (2001). *Initial Public Offerings*. IN CHEW, D.H. J. (Ed.) *The New Corporate Finance*. *Where Theory Meets Practice*. 3rd ed. McGraw Hill-Irwin.

Jordan, J., Lowe, J. and Taylor, P. (1998). 'Strategy and Financial Policy in UK Small Firms', *Journal of Business Finance and Accounting,* **25** (1-2), pp. 1-27.

J. Chris Leach & Ronald W. Melicher (2010). *Entrepreneurial Finance*, 4th Edition. South-Western, Cengage Learning.

Kato, M. P. and Kratzer, J. (2013). Empowering Women through Microfinance: Evidence from Tanzania. *ACRN Journal of Entrepreneurship Perspectives Vol. 2, Issue 1,* p. 31-59, Feb. 2013.

Kessy, S. and Temu, S.S. (2009). Impact on Training on of Micro and Small Enterprises Served by Microfinance Institutions in Tanzania: *Research Journal of Business Management*.

Kessy, S.S.A. and Urio, F.M. (2006). The Contribution of Microfinance Institutions to Poverty Reduction in Tanzania. REPOA Research Report 06.3. Mkuki na Nyota Publishers, Dar es Salaam, Tanzania.

Klapper, L., Laeven, L., & Rajan, R. (2010). Entry Regulation as a Barrier to Entrepreneurship. *Journal of Financial Economics*, **82(3**), 591-623.

Kilolo District Council, Social Economic Profile, 2013. Iringa Region.

Kira, A.R, He, Z. (2012). The Impact of Firm Characteristics in Access of Financing by Small and Medium-sized Enterprises in Tanzania. *International Journal of Business and Management; Volume 7,* No.24; 2012.ISSN 1833-3850, E-ISSN 1833-8119. Published by Canadian Center of Science and Education.

Mbura O. K. (2013). Entrepreneurship Development and Management in Tanzania. Theoretical and Practical Reflections. NBAA, Dar es Salaam.

Mori,N & E. Richard, A. Isaack, and D. Olomi (2009). "Access to finance for SMEs in Tanzania," in African Entrepreneurship and Small Business Development: Context and Process, vol. 1, D. Olomi, Ed. Dar es Salaam: Otme Company Ltd., 2009.

Myers, Stewart C. (1984). Presidential address. " The Capital structure Puzzle". *Journal of Finance* **39,** no.3; 575-592.

Myers, Stewart C., and Nicholas S. Majluf (1984). Corporate Financing and Investment Decisions When Firms Have Information the Investors Do not Have." *Journal of Financing Economics* 13, no.2; 187-221.

Shafeghat, A., H.Darvish, A.A. Namdar and R. Gholami (2008). Pattern of Rural Entrepreneurship in Economic and Social Development. Paper Collection of National Conference on Sustainable Rural Development with Emphasis on Agriculture, pp: 275-276.

Mkpado, M. & Arene, C. J. (2007). Effects of democratization of group administration on the sustainability of agricultural microcredit groups in Nigeria. *International Journal of Rural Studies*, **14 (2)**, 1-9.

Murray, F.Z. and Goyal, V.K. (2003). "Testing the pecking order theory of capital structure", EFA0157: AFA 2001, New Orleans, LA, available at: http://ssrn.com/abstract=243138.

Ministry of Finance, Tanzania (MOF 2012; 2013), Speech by the Minister for Finance Hon. Dr. William

Augustao Mgimwa (Mp). Introducing to the National Assembly, the Estimates of Government Revenue and Expenditure for the Fiscal Year 2012/2013 and 2013/2014.

Nassr,I.K., and Wehinger,G.(2015); Opportunities and limitations of public equity markets for SMEs; OECD Journal: *Financial Market Trends*. Volume 2015/1.

Nachman, D.C., Noe, T.H. (1994). Optimal Design of Securities under Asymmetric Information. *Review of Financial Studies* **7,** 1-44.

Nghiem, H.S. (2009). Analysing the Effectiveness of Microfinance in Vietnam: A Conceptual Framework, JEL classification - O12, P34, R29 School of Economics. The University of Queensland.

Nyamsogoro, Ganka Daniel. (2010). Financial sustainability of rural microfinance institutions (MFIs) in Tanzania. University of Greenwich.

Olomi, D. R. (Ed.) (2009). African entrepreneurship and small business development: context and process. Otme; Dar es Salaam.

Olomi, D.R (2005). *Enhancing SMEs Contribution to Employment and Poverty Reduction within National Poverty Reduction Strategy*. Workshop Paper on Enhancing SME's Access to Government Procurement and Food Standards by Association of Tanzanian Employers (Ate); Regency Park Hotel, Dar-Es-salaam, 30th September. Retrieved from www.tanzaniagetway.org

Osei-Assibey, E. Bokpin, G.A. Twerefou, K.D (2012): "Microenterprise Financing Preference; Testing POH within the context of Ghana's rural financial market. Journal of Economic Studies, Vol.39 No.1, 2012, pp 84-105. Emerald Group Publishing Limited 0144-3585.

Paul, S., Whittam, G. and Wyper, J. (2007), "The pecking order hypothesis: does it apply to start-up firms?". *Journal of Small Business and Enterprise Development*, Vol. 14 No. 1, pp. 8-21.

Peter, B. K. (2001). *Impact of Credit on Women-Operated Microenterprises in UASIN GISHU district*, Eldoret, Kenya. In P. O. Alila & P. O. Pedersen (Eds). (2001). Negotiating Social Space: East African Microenterprises. ILO, Geneva, database. [Online] Available: http://books.google.com.my/book.

Poutziouris, P., Chittenden, F. and Michaelas, N. (1998). *The Financial Affairs of Private Companies*, Tilney Fund Management, Liverpool.

Randhawa, Bikki K, & Gallardo, Joselito. (2003). Microfinance regulation in Tanzania: Implications for development and performance of the industry: World Bank.

REPOA (2008*),"The role of small Business in Poverty Alleviation: The case of Dar-es-Salaam, Tanzania"*. Mkuki and Nyota Publishers, Dar-es-Salaam, Tanzania: 9pp.

Rubambey, G. (2005). Policy, Regulatory and Supervisory Environment for Microfinance in Tanzania. The Essays on Regulation and Supervision, No.15, CGAP.

Rutherford, Stuart (1999). "The Poor and Their Money – An Essay about Financial Services for Poor People", Institute of Development Policy and Management, Manchester, 1999.

Saunders, M., Lewis, P. et al. (2007). *Research Methods for Business Students*, 3rd Edition. Prentice Hall, Pearson Education.

Scherr, F.C., Sugrue, T.F. and Ward, J.B., 1990, 'Financing the Small Firm Start-Up: Determinants of Debt Use', *Journal of Small Business Finance*, 3 (2), pp. 17-36.

Schreiner, M. & Leon, J. (2001). "Microfinance for Microenterprise: A Source Book for Donors", Savings and Development, 2002, Vol. 26(4).

Stewart, Ruth, van Rooyen, Carina, Dickson, Kelly, Majoro, Mabolaeng, & de Wet, Thea. (2010).

What is the impact of microfinance on poor people?: a systematic review of evidence from sub-Saharan Africa.

Tundui, H.P. (2012). *Gender and Small Business Growth in Tanzania*: The Role of Habitus: University of Groningen, the Netherlands, 161 pp.

Qin, X., & Ndiege, B. O. (2013). Role of Financial Development in Economic Growth: Evidence from Savings and Credits Cooperative Societies in Tanzania. *International Journal of Financial Research,* **4(2),** 115-125. http://dx.doi.org/10.5430/ijfr.v4n2p115.

Wangwe, S. (2004). Innovation in Rural Finance in Tanzania. Paper prepared for The Third Annual Conference on Microfinance held from 15th to 17th March 2004 at the AICC, Arusha, Tanzania.

Wit, G. & Van Winden, F. (1989). An Empirical Analysis of Self-Employment in the Netherlands. *Small Business Economics,* **1 (4),** 263-272.

Wyer, P., Ekanem, I., North, D. and Deakins, D. (2007), "Literature review for the small business service: the impact of perceived access to finance difficulties on the demand for external finance: final report", Centre for Enterprise and Economic Development Research, Middlesex University, London, available at: www.berr.gov.uk/files/file40922.pdf.

Wangwe, Samuel, & Lwakatare, M. (2004). Innovation in rural finance in Tanzania. Paper presented at the Third Annual Conference on Microfinance held from 15th to 17th March.

A Perfect World? Risks and Threats in the Information Society

Gustavo A. Masera★
& Javier Ulises Ortiz★★

About the Authors

★Gustavo A. Masera, Ph.D. in History, National University of Córdoba. Multidisciplinary Institute of Contemporary Social Studies (IMESC-IDEHESI), Faculty of Philosophy and Letters, National University of Cuyo, (Mendoza), Argentina. Email for correspondence: gustavo.masera@gmail.com

★★ Javier Ulises Ortiz, Ph.D. in Political Science, Universidad del Salvador. Professor and research at National Defense University, (Buenos Aires), Argentina. Email for correspondence: juo425@yahoo.com.ar

ABSTRACT

The document provides an approximation to the issue of global systemic risks in the information society and, in particular, those having an impact on critical infrastructures. In order to do this, the experience of mature developed countries is taken into account from the perspective of agencies and authors specialized in these issues. The discussion will allow not only for the comprehension of the risks associated to the development of these infrastructures but also value governance as a participative mechanism for risk management. The result of this work will have useful features as support for decision making and the formulation of domestic and foreign public policies, and especially support the debate being born in Latin America.

Keyword: Risks, Critical Infrastructures, Information society, Latin America

"New and terrifying events are happening as we take our first steps into the 21st Century", OECD, Emerging systemic risks in the 21st Century.

CITATION:

Gustavo A. Masera & Javier Ulises Ortiz (2018). "A Perfect world? Risks and Threats in the Information Society." *Inter. J. Res. Methodol. Soc. Sci.*, Vol., 4, No. 1: pp. 36-49. (Jan. – Mar. 2018); ISSN: 2415-0371.

1.0 THE INFORMATION SOCIETY

The shared conception in the academic world is that scientific-technological changes which have taken place over the last decades of the 20th Century constitute a theoretical-conceptual challenge, an opportunity field for economic evolution and a set of threats and risks for society as a system (Bischoff, 2008).

The discussion posited in the work will allow to: (i) To sustain the elaboration of appropriate strategies for the insertion of Latin America in the international community known as “information society” (IS), in a more theoretical plane; (ii) To plan the path of growth of critical infrastructures bearing in mind the risks that goes with them; and (iii) To implement participation mechanisms for the management of the IS and of critical infrastructures based on sustainable development and governance (Bevir, 2007).

In a more operative plane, to formulate bases for public policies aimed at the management and regulation of risks. It is in this frame that an approximation is done to the issues of the information society. In particular, the risks of critical infrastructures are considered, taking into account the experiences of maturely industrialized countries, think-tanks perceptions and authors specialized in these matters (IRGC, 2006). A realistic analysis identifies the potentialities, limitations and dangers involved in the development of the IS. The biggest gain will possible be a comprehension of the entirety of positive and negative aspects and, above all, its implications for the IS.

The first thing to consider is that the international contemporary system has undergone enormous transformations over the last few decades, with many of these changes coming over the last few years. Amongst these transformations, we can identify:

The unification of the world economy through a multidimensional and complex process of globalization in an international setting which emphasizes existing asymmetries and cannot but lead to an unequal development of the different actors. In this new world we find that centrifuge forces of dispersion, fragmentation and crisis (with important geographical transference of industrial installations and introduction of technological changes) live together with centripetal forces of intertwining and interdependence between different regions of the world (having as principal vector the interconnection of the infrastructures of information, energy and transport). This process is accompanied by a growing internalization of diverse actors such as companies, financial institutions, etc. (Masera, 2010; Nye and Donahue, 2000).

The centrality of the socio-technological paradigm endlessly multiplies new models of scientific-technological production. These new models simultaneously affect the pre-existent means of production (from agriculture to traditional industries, such as the automobile sector) and give origin to new areas of products and services (for instance the production and distribution of alternative sources of energy). These new models are generally localized in systems of high specialization territorial innovation enhanced by digital infrastructures and communication networks of universal access (Mansell, 2009; Bernal-Meza and Masera, 2007).

A renewed tendency to the formation of regional spaces of commerce and investments with open agendas for the discussion of a variety of issues and the degree of depth for negotiations within a regionalization dynamics of political, economic and cultural systems. These integrated spaces tend to cut across national borders and jurisdictional spaces and are strengthened by infrastructure networks which facilitate contact and exchanges (Masera, 2010).

The emergence of systemic risks with global dimension fed by the same factors which make the abovementioned points possible: economic unification, technological intensity and regionalization. Some recent examples that can be cited are: the 2008/09 international financial crisis, economic recession, aviary and a flu pandemics, food crisis, the commercial negotiations of the World Trade Organization (WTO) getting stuck, problems with supply of gas from Russia to Europe, power blackouts of great dimensions in the US and Europe, consequences of climate change (Lippert, 2016), etc. These risks can affect the socio-political stability and the economic prosperity of many nations. In addition they can provoke other changes that can have lasting effects on the distribution of power among countries and regions, and even catastrophic impacts on the most vulnerable societies (Masera, 2008).

The global society of the beginning of the 21st Century is characterized by the massive use of technology information and of communications (ICT).This insertion of the ICT is changing the modes of governing, producing, socializing, and offer and access to goods and services. In advanced industrial societies, this evolution has given place to the birth of the IS (Webster, 2006;

2002). Even if there is not universally accepted definition, it can be stated that the information society is a post-industrial society (or knowledge society) where the generation, circulation, manipulation and use of information processed through computers and digital communications is an essential component to social, political and economic activities (Hadkiewicz, W. and P. Gawowicz, 2013; Peres, Hilbert, 2009; Comisión Europea, 2003).

2.0 SYSTEMIC RISKS: GENERAL CONCEPTS

Talking about risks that can affect society as a whole, the issue of global emerging risks of a systemic nature cannot be ignored (IRGC, 2010).

The international setting has seen the emergence of some hot issues since the end of the 90s such as the evidence of climatic change and its effects on natural resources, accelerated population growth and migratory flows, the decay of public services in less developed regions and social tensions in societies which receive immigration. It has also been established that the technologies used to solve some of these problems (for instance, new vaccines or seeds) or to promote economic development (with ICT as a priority) are not risk-free. More importantly there is a maturity in the understanding (as it has been partially experienced) that these risks have the capacity to cause strong impacts on societies, with even the potential for destructuring them.

The 9/11 attacks in USA and other events like the massive blackout in big urban agglomerations, have demonstrated that the unlimited security in a complex world in impossible (Ortiz, 2008). The cyberspace is the new "Athena´s Camp" for the development of new conflicts, many of "asymmetrical" type (Thomas, 2005). In the developed countries the concepts of cyberwarfare, information warfare and information operations appear as new governmental cybersecurity / cyberdefense doctrines) (Castells, 2002).

According to the sociologist Manuel Castells the 9/11 was the beginning of the first world war of the XXI century, the "net war" that attempt to "impose their objectives by using the only efficient weapon in its technological and military inferiority situation."

In May, 2007 the Estonia computing system was attacked through Internet by a half of millions of computers. Estonia was paralyzed during some weeks y needed the NATO help. James Appathurai, a NATO spokesman, said in relation with the attack: "the XXI century is not of tanks and artillery" and after the meeting of the NATO chiefs of June 14th, 2007, he summarized that: "everybody agreed that it is indispensable to improve the protection capability of the computing systems of critical importance" (BBC, 2007).

As an answer to these attacks, a new concept in defense and security matters appears: the Protection of the Critical Information Infrastructure (PCII) to reach bigger possibilities of endurance against these attacks and accidents. It is necessary to identify and secure the CII to avoid a new "Mutual Assured Destruction" for this new age. The cooperation hemispheric agenda for the regional security and defense starts to incorporate these definitions so that the member countries develop their own concepts. The telecommunication (optic fiber, digitalization, computing) are the technologic infrastructure of the globalization that make possible the strategic decision making in real time at a global scale (Ortiz, 2008).

Since the late '90s the nations acquire new forms like the "Digital State" (Keyworth, 1998) or "Net-State" (Castells, 1998), and their sovereignties are enlarged. Whether by attack or by accident (blackout in megalopolis), the risks of not having prevision systems, early warnings and fast answer based on the emergency plans, can be devastating

This danger affects societies with advanced economies and those in process of development though with different characteristics. What is surely different is the faculty of resistance, resilience and elasticity shown in these three kinds of societies regarding risks. As a result, the fundamental issue for the governance of societies, specifically in Latin America, is that of institutions, decision-making processes and the instruments to face these inevitable risks. To participate in the 21st Century world means to be exposed to the risks already mentioned and the biggest risk is to ignore this situation.

As examples, we can point out cases that allow us to hypothesize other possible crisis of a bigger, more dramatic dimension:

- "Mad cow" disease (especially in Europe and particularly in the U.K.) which involved costs, economic effects, changes in consumption and commerce patterns.

- The aviary flu, which had an effect on international tourism, on consumption of foods, on human health and even on education and scientific investigation programs.

- A strong conflict between new OGM products and traditional harvests (with advances in biology, genomic structures, etc.) This generates social acceptance effects, changes in genetic diversity, effects on world commerce and on the monopoly of patents.

A set of new applications and materials based on nanotechnology and its applications in food products and cosmetics which has effects on health, destruction of markets with seasonal changes of products and the strengthening of patents monopoly.

The possible faults in ICT systems used uniformly around the planet (e.g. operative systems, ofimatics programs), and the possibility of cyber-attacks on the whole of society, like in Estonia in 2007 and Georgia in 2008 (Ortiz, 2012; Saadawi. and Jordan 2011) and in the European Union in 2017.

The fragility of the security and safety systems in critical installations as in nuclear plants, with Chernobyl symbolically, showing the borders of modernity. The acceleration of changes and contagion mechanisms in an interdependent world economy imposed by the velocity, connectivity and intangibility in the different infrastructures but, especially in the financial one (IRGC, 2016).

We can add natural disasters to the list (like earthquakes and tsunamis) which have affected all civilizations in different historical epochs but, whose effects could be more disastrous nowadays due to the difficult continuity of life after the destruction of basic infrastructures as shown by the case of New Orleans after Hurricane Katrina in 2005 (Moynihan, 2009). There is also the case of disasters caused by human action on nature such as intentional fires. All this gives the impression that society is at moments on the edge of losing control when faced with important risks, threats, disasters and non-conventional crisis (IRGC, 2009).

Some of the questions that can be asked are:

Do we know the risks to which our societies are exposed?

Do we know the fragility of our society regarding these risks?

Are we prepared to face them? Do we know how to prevent, intervene and/or mitigate the consequences?

Do we know what capacities we have for countering them, be them organizational, technological or human? Have we planned what capacities we will have to develop over the next few years in order to be prepared for future risks? Who has to take which decisions? Which is the role of public authorities and which is the role of infrastructure operators? And, what are the voice and the role of civil society?

3.0 SYSTEMIC RISKS: DEBATE AND APPROACH

In a first approach, we privilege the identification of the forces generating systemic changes. These forces permit the formation of risks in 4 fundamental contexts: demography, environment, technology and socioeconomic structures. It cannot only be considered that new changes may arise but also, that the alteration of conditions may imply the spreading of negative consequences. This would imply the transformation of the channels through which "accidents" spread and might even alter the kind of social responses (OECD, 2003).

Of course, there are technical and non-technical dimensions to the problem of risks. The technical elements may, up to a point (in the absence of local availability), be acquired in the world market, taking as an example the best practices carried out in advanced countries. Technical elements though can neither be improvised nor copy from other experiences because they depend intrinsically on the risk environment and the society being affected; also, its appreciation depends on the psychological and societal experience of the risk. Besides, it depends on the perception of the possibility of it happening, the magnitude of its possible impacts and the importance of preventing negative situations which are by nature fortuitous and not predictable.

Let us remember that Thompson started by distinguishing (1990) between real risks (what may occur, with negative consequences that can happen with a statistically known probability, quakes, plane crashes), observed risks (what can be deducted from models or studies like the possible effects of an epidemics) and, perceived risks (subjective judgements in the absence of models or previous knowledge) like the effect of future nano-technological products. There are differences between risk and risk perception (Campbell, 2007).

It is possible to defend the following idea: Even if models and statistics may help, the perception of risk is always a persona/social judgement and it does not always coincides with outer reality. This is specially the case of complex systems like infrastructures. The point being that nobody really knows for sure how these will function in abnormal situations (at least the very significant ones like large integrated electrical systems, the Internet, etc.). There is not and there cannot be any certitude about the probability and the effects of risks on complex systems. The point about infrastructures is above all to understand their tendency to grow and develop. This is the prevailing direction: growing complexity (Pierre and Peters, 2005).

Therefore, the concept of risk is relative: the same effects can be rated in diverse manners by different societies and people. Moreover, there are risks that do not affect all citizens in the same way – and there can be winners and losers with different risks and options regarding action (Lechte, 2003).

So, we have to talk about the acceptability of risks, including the activities for preventing, monitoring, acting against and mitigating them.

In any case, assessment, management and evaluation of alternatives of risks is fundamental since there are critical systems whose failure may produce the collapse of society, just as it has happened before in the history of mankind.

Some factors that must be considered when talking about risks are:

- Uncertainty as regards the probability of occurrence, for instance, what is the probability of a catastrophic failure in an infrastructure in 10 years?

- Uncertainty about the severity of the impact of a catastrophic failure, for instance, causing a huge blackout (Gheorghe et al., 2005).

- Possible victims and damages (to health, environment, properties) and their knowledge about the existence of risks, dangers, etc.

- Whether it is possible to reverse negative effects.

- Compensation for being exposed to risk, or the existence of the option to choose not to be exposed.

- Benefits and dangers (and costs) for the different actors.

4.0 RISKS IN THE INFORMATION SOCIETY

Within the general realm of contemporary risks, it is necessary to consider the specific risks arising in the context of the IS.

The concept IS makes reference to a new paradigm whose orienting criteria show a new type of organization still being built (CEPAL, 2003). The IS, is the result of the action of technological systems enabled and empowered by the new information and communication technologies together with the progressive processes of digitalization. The fast planet-wide spreading of this expression has been given impulse by the action of various international organisms together with the publicity owed to some futurologists and enterprise management gurus (Masera, 2010).

As objective data about it coming into the scene, there are numerous initiatives like summit meetings, seminars, observatory bodies, publications, forums and entities which have taken place recently organized with the purpose of analysing the transition towards this digital society. An illustration of this took place in 2006 with May 17 being declared IS's world day by the UN. Arising from what was agreed upon in Tunisia 2005, "World Summit IS", resolution A/RES/60/252 of the General Assembly was approved with the objective of commemorating the 140th birthday of the International Telecommunication Union founded in 1865, which was the first worldwide intergovernmental organization (ITU, 2006).

Just over a decade ago, Ulrich Beck (1992) introduced and made popular the concept of "risk society". Beck points out problems and criticises the limitations of the current society but does not go as far as proposing a way of managing these problems. One may share his call for a participative democracy, but, the idea that the current technology has created new risks can be rescued as well as that of qualitatively different dangers, when compared to those of the past (Mythen, 2004).

It can be averred beyond doubt that no system is perfect. In the IS there exist risks, threats and vulnerabilities -and points close to errors, failures, security, etc- unbeknownst before. (Masera, 2010). The ICTs present solutions which appear dazzling for their effects of advanced modernity. Actually, its modernity resides both in knowing how to make use of its functionalities as well as knowing how to manage its limitations. On the other hand, as many of current technological risks do not respect national frontiers, there also arise problems relative to regional or international coordination policies regarding them.

The challenge is that the matter of weak points which appear when connecting the "whole" to Internet must be confronted rapidly, as they might affect the future of commerce, government, hospitals, banks, energy, etc. The so-called 'Information and Communication Technology' (ITC), which is the basic support for the IS, is the result of the interconnection of telephony systems to the Internet, and satellite communications. It can be observed how other systems, as the GPS and all means of information come together through this great IIC. This dynamic process of convergence of the ICT technological sectors together with the formation of a network of networks on the one hand gives place to the development of all "e-" services: e-commerce, e-government, e-health, etc.; while though on the other hand it inevitably sows fragilities. Hardware and software and adjoining programs are not, and perhaps will never be, fool- and error-proof.

Also, as many of the current technological risks do not respect national frontiers, there arise problems too, relative to regional or international coordination policies. The International Internet Connectivity (IIC), as defined for instance in the 'World Summit of the Information Society' in Tunis 2005, coordination and those of the IS as a consequence are by nature local, regional and global realities at the same time.

The trustworthiness of infrastructures, complex socio-technical systems and the trust that the citizen and society can deposit on them are at the core of the matter of IS governance. The advantages brought by the massive and ubiquitous use of ICTs may, concurrently become significant threats due to their security problems, at levels and extensions which are difficult to foresee as it is hard to predict the evolution of technologies and their future uses.

What is true is that, be it because of accidents with risk for physical or logical effects, destruction or decay of integrity or data disposability, failures or attacks affecting confidentiality, or the privacy of users the information society is destined to suffer continuous security problems. Some of these problems will be accepted as minor and tolerable while others will cause major damage to people or companies, and, of course there will always be the possibility of exceptional events, though not impossible, with potentially catastrophic effects. This is the issue to be explored in this article.

From this point of view, the ITC is a critical infrastructure of the IS. Horrible failures in its functioning leading to the collapse of the technical base of the IS for a certain period will result in the loss of data transmission and access to sources of information thus affecting other critical infrastructures (water, transport, electricity, logistics) and could within days lead to the collapse of society.

The risk, concretely refers to the possibility of damage or harm in a given sector, for instance, the information and communication infrastructure together with the extension of this damage to all other infrastructures that depend upon the IS.

There are at least two perspectives about the risk: an objective one that can be measured and a subjective one, not only at personal level but also having social components, and referring to the perception of danger. The risk then does not refer exclusively to specific damage; on the contrary, it takes into account the consequences on society as a whole.

According to the International Risk Governance Council (2009; 2006) the risk of critical infrastructures must be thought about from the point of view of the consequences more or less certain from events or activities and from their potential impact regarding what is valuable be it for society as a whole, be for a group or for a single individual. As an example this can involve concrete things such as natural goods, the environment as a whole, human health, natural resources, etc. as well more abstract elements like social and economic stability, privacy, etc.

It must be born in mind that the inherent characteristic of risks is uncertainty and therefore, since risks are uncertain and contingent to many undetermined factors, it is difficult to foresee them in an analytical way or be based on statistical values. One of the main tasks is to imagine future scenarios, that is, the possible shapes these risks can adopt through a systematic study of possible future situations with the goal of better defining defensive and palliative strategies.

5.0 RISKS OF CRITICAL INFRASTRUCTURES AND DEVELOPING COUNTRIES

According to certain initiatives and to the experience of the most advanced societies that have gone deeper on this issues over the last decade, risks can be defined bearing in mind the 5 main systems of critical infrastructures: Supply of energy; Supply of drinking water; Treatment of waste; Transport; Information and communication infrastructure (that is digitally based including Internet, used for managing, monitoring and controlling the other infrastructures).

The “Critical Infrastructures” are a network of interdependent systems at great scale and imply complex physical transnational distributions associated to technologies and cybernetic networks, product of the interconnection with information and communication systems (Gheorghe et al., 2013; 2005). Here we can consider three main aspects:

a. Their function is to produce a continuous and universal flow of basic services which are essential for the economic and social development. In other words they are elements which have to be at everyone’s disposal at all times. The user is not worried about the complexity present behind his access to the service; he wants just to be connected and that the service be available.

b. They tend not to be the possession of only one owner (public or private), operator or regulator or user, and have diverse functioning logics. After the process of opening of markets which took place in the world as from the 1980s, infrastructures were no longer seen as natural monopolies. The technological development has been used for opening up market infrastructures (see for instance, the process called “unbundling” of the European electricity system REF). Ideally, no single operator controls the infrastructure and when the systems are interconnected across borders, the very same national regulatory bodies see their powers being clipped.

c. They have been designed so as to satisfy basic social needs, but technological and organizational changes have raised their complexity and are bound by internal and external risks due to intentional or accidental failures. When these failures do occur, they tend to propagate and exceed the structural, functional and territorial limits of each single system.

From the point of view of developing countries, some of the main elements to be considered as regards critical infrastructures are the following:

(i) The progress of paradox: more development leads to more fragility. Infrastructures are critical

The development of the economy has been determined in the 20th Century and at the beginning of the 21st Century by the growing incorporation of technologies improving productivity and innovation. This affected the industrial production of goods and services and the infrastructures supporting them. Their evolution is characterized by the massive integration of technologies, mainly information and communication and by the mutual integration and interdependence (for example, between transport systems, energy, financial and communications.).

In order to be competitive at the international level, it is essential to count with adequate means of transport, telematic access to markets, dependable energy supply, etc. Not to mention the same chains of supply necessary for production (e.g. fertilizers).

But, as it has been proved in regions like North America, Europe and Japan, the increase in infrastructures and their technological content makes them more fragile to all kinds of possible threats, be it with regards to loss of important capitals invested or to the damages that a decrease or perturbation of the infrastructural services or, the very destruction of the infrastructures and what this can provoke: good examples of these are natural risks, technical failures of systems, human errors in complex situations and probable fraudulent attacks such as terrorism, organized crime, radical groups, war (Ortiz, 2012; NIPP; 2009).

It is for this reason that developing countries must worry not only about the investment in the growth of their capacities in infrastructures but also in their protection and in recognizing their criticality and vulnerability.

Possible accidents, like Hurricane Katrina or an earthquake, or international conflict situations like the one in Estonia, Georgia or the Balkans which will affect the critical infrastructures might cause vital damage to society.

(ii) The management of infrastructures and their critical aspects requires a new kind of collaboration between public and private actors: system of governance

Over the last few years there has been a deep transformation in the management of infrastructures: from public services in the hands of the State,or in the hands of government bodies, generally in a situation of monopoly, to a kind of enterprise management, mainly in private hands and with a tendency to abide by economic market rules.

Infrastructures managers are therefore the main investors in these systems and resources, and they do so with a double objective: to satisfy the needs of social and economic actors served by this infrastructure (like in the case of drinking water, electric energy, etc.), but also to guarantee the continuity and the quality of services since the good functioning of society depends on this condition.

These guarantees are usually detailed in the contracts of allocation of the service. But, in an open market with many competitors and rapid technological evolution, and close and accented interrelations between infrastructures, it is not always evident what is the necessary level of protection and security. To leave such a delicate issue up to the will and decision of companies does not seem effective since investments in security and the trustworthiness of the service normally do not have an immediate return. In the case of several actors competing, it may occur that the different types of investment degenerates the balance of the markets. Examples of this are problems such as “free riders” and “moral hazards”.

In these cases the legitimate enterprise interests may be in conflict with society’s needs (even without considering enterprise fraud). In order to solve this, society needs a new approach to govern situations of public relevance but managed by the private sector.

The solution to this problem must consist of the implementation of new governance models with the participation of all actors involved. This also implies shared decision-making. This may be linked to the definition of standards (technical and belonging to processes), control procedures etc. The great challenge is to make compatible and synergic the market mechanisms and the strategies for the security of the state.

(iii) Critical infrastructure as crucial pivot for innovation

Infrastructures can be innovation engines from two different points of view. First, infrastructures constantly demand new products and services and also need new internal processes and in relation to his productive chains (suppliers, clients, etc.). This demand is materialized in relation to the elements necessary for the continuity and efficiency of operations, but also in relation to the security and dependability in the supply of services.

On the one hand, infrastructure operators have to invest in a continuous way in technologies in order to be able to produce in a profitable manner, seeking to keep being competitive in the market; and in order to respect different standards and norms (such as environmental for transportation and energy). They have to protect their infrastructures from various possible dangers and correct possible problems caused by the technologies (like software mistakes). On the other hand, if the infrastructures are qualified and competent, they may act as innovation catalysers offering new and better services at a better price.

This last point is particularly true in the case of information infrastructures with the developing of the IS. It is evident that Internet and mobile communications allow the growth of new industrial, commercial and public services capacities (e.g. e-government, e-health, e-commerce, etc.)

(iv) Critical infrastructures as comparative advantages: human resources

Infrastructures are designed, operated, maintained and used by qualified staff. The training of these persons is crucial for the efficiency and competence that can be obtained within the infrastructures. As many of the subjects involved are new, not always the academic institutions are prepared for satisfying the requirements of industry. No the least, many issues are multi-disciplinary, cutting across the vertical lines typical of the subjects offered in universities.

The most developed countries are already investing in the formation of technicians and professionals who will have responsibility over those infrastructures. This requirement not only refers to technical staff, but there is a need for lawyers, economists, etc.

6.0 GOVERNANCE AS A RESPONSE

There is one aspect of the problem on which all decision-makers agree upon: risk must be managed and controlled, but how? Living with risk confronts contemporary societies with questions of a political character. Here is where governance comes in.

Governance can be understood as the structure and processes for a collective decision involving government and non-government actors. (Nye and Donahue, 2000).

The EU has defined in its white book about principles for good governance. They must be applied to the specific case of identification, evaluation, management and communication of risks. Amongst other principles, governance presupposes the participation of all political, economic, social and scientific-academic actors in consulting aspects, starting from a bigger sense of responsibility in the decision making and the participation of citizens in the elaboration and application of policies .It requires transparency in the communication of decisions and efficacy regarding the taking of decisions in scale and at the appropriate moment.

Risk governance refers to the capacity of organizations and people to face inevitable risks (Aven and Renn, 2010). As a management tool it refers to processes oriented to the taking of decisions where comprehensive, whole solutions can be defined and found for all relevant actors with interests and who are involved in the particular subject (stakeholders). It is in this sense that governance is a process of participation and consulting useful for managing complexity. The idea is that only the management of complexity can reduce complexity.

A very important example of the new risks and of risk management in a regional context through governance mechanisms is given by the crisis in the interconnection European electrical system in 2004.

The teaching from this experience reveals that in the contemporary world there is a continuous widening of the threat spectrum and that it must be faced by all infrastructures, including critical ones; this includes a variety of natural occurrences, technical failures and accidental or intentional human actions which challenge the stability of networks and the safety of some services which should never be interrupted.

The point is that when the whole functioning depends on one complex technological system, one single failure may have catastrophic consequences (Renn, 2008).

An adequate risk management in critical infrastructures has the advantage of:

- Reducing the growing interdependency in the negative aspects of crisis transmission.
- Promote the management of the demand and the adjustment of priorities.
- Reduce the times for restoring the system after a failure and allows for the maintenance of critical social services when facing a failure in the system.

7.0 HEMISPHERIC STRATEGIES FOR THE PROTECTION OF CRITICAL INFRASTRUCTURES (PCI) AND THE CYBER SECURITY IN THE AMERICAS

Different agreements celebrated during the last years by the countries of the Organization of American States (OAS) settle the basis for the strategies of PCI and cyber security:

- The Hemisphere faces an increasingly diverse and complex set of threats and "challenges for the states, societies and people" and there are "particular strategic context of each sub-region in the Hemisphere", the civil and civil-military cooperation in the fields of Defense and Security is necessary. V Ministers of Defense of America Conference (Santiago de Chile, November 2002).

- The new concept of security in the Hemisphere is multidimensional in scope, and includes traditional and new threats, concerns, and other challenges to the security of the states of the Hemisphere, incorporating the priorities of each State. The security of States of the Hemisphere is affected, in different ways, by traditional threats and the following new threats, concerns, and other challenges of a diverse nature and it includes "attacks to cyber security" and consider "new terrorist threats, whatever their origin or motivation, such as threats to cyber security, biological terrorism, and threats to critical infrastructure" OAS, Declaration on Security in the Americas (OAS, 2003).

- An Inter-American Strategy to Combat Threats to Cyber security. "A multidimensional and multidisciplinary approach to create a culture of cyber security" to protect de infrastructure of the telecommunications and give responsibilities to the Inter-American Committee against Terrorism (CICTE) for a formation of an Inter-American Alert, Watch, and Warning Network to Rapidly Disseminate Cyber Security Information and Respond to Crises and creation of the hemispheric network of a Computer Security Incident Response Teams (CSIRTs).

-Inter-American Telecommunications Commission (CITEL): for the identification and Adoption of Technical Standards for a Secure Internet Architecture (OAS, 2004).

This important document is a key to consider firstly a common vision of the Critical Infrastructures in the Americas.

-Hemispheric definition of Critical Infrastructure (CI) "refers, among others, to those facilities, systems, and networks, and physical or virtual IT services and equipment, the disabling or destruction of which would have a severe impact on populations, public health, security, economic activity, the environment, democratic governance, or the ability of the government of a Member State to operate effectively" (OAS, 2007). This important document is a key to consider firstly a common vision of the CI in the Americas.

-Trends of Cybersecurity in Latin America and the Caribbean. These first general mapping of the cybersecurity in Latin America show that the region features a youthful population craving more online access. The recent situation regarding cyber threats in Latin America and the Caribbean shows that users are suffering the impact of threats that are prone to global level and others specific to each region. As an aggravating factor of this challenge, Latin America and the Caribbean have the fastest growing population of Internet users of the world, with an increase of 12 percent during the last year. This report identified the main trends that impact the region: the weak capabilities of law enforcement agencies, and the complex forms and scale of crime online (OAS, 2014).

8.0 SOME LESSONS FOR LATIN AMERICA (POINTS TO BEAR IN MIND)

Latin America must have an active participation in the construction of the IS. The aim is to "incorporate the paradigm of the IS in the development agenda"; even more so when there is an attempt to re-launch the program on the development of infrastructures in the plane of intra-South American relations. A true conception of regional integration cannot leave behind the question of infrastructures and that of the associated risks to this process. It cannot be forgotten that new threats and vulnerabilities appear from the crossing between new information technologies and the communication with infrastructures (Girard and Perini, 2013).

These are some of the ideas that synthesize the perspective of this work:

(1) The diverse issues that appear from the construction of the IS will internally influence the definition of the quality of life, wellbeing and growth projection of citizens; internationally, it will influence the definition of interests in foreign policy, in the regional integration processes and the general behaviour of the region.

(2) In the IS, the objective should be to develop means and capabilities, and to have better infrastructure with an adequate use of technology- But it should be always taken into account that this carries along fragility, and so the potential benefit has a strong interconnection with the potential negative effects. The challenge of a good risk management rests then on the utilization of the benefit of information and communication technologies while minimizing the negative consequences associated with the risks.

(3) It is absolutely necessary to have risk management of regional or global character, especially of those that have a high level of impact on health, security, the environment, the economy or society- In addition, these should be evaluated and managed through widely participative consulting mechanisms. This means that the development-risk duality needs to be analyzed by the diverse interested and committed publics (stakeholders), and also it must be included as a variable by formulators of policies and managed by political authorities.

(4) Infrastructures are operated by companies (private or public, but with a business organization and a search for profit). Therefore, there is a need for governance as a management tool at the different levels of territoriality and particularly applied to the integrated management of multiple actors, multiple interests and perspectives. Risk governance is the only modality which allows the convergence of the interests of companies with the requirements and integral demands of a society.

(5) The specific risk in critical infrastructures has to be considered as object of analysis. Security in the supply of the service and the impact that might occur if there were any long interruption of the services, should constitute a high level priority for legislation, coordination of policies, planning and the evaluation of scenarios.

(6) The global risks are not contained within national borders and therefore, cannot be managed through actions or policies by just one sector or isolated government. Thus, risk governance, especially those of regionally interconnected critical infrastructures, requires a special coordination among the countries that may be involved.

(7) From the perspective of the socio-economic implications, infrastructures in their fast evolution "consume" products and services at the same time that they show the way to new social and economic capacities. This situation defines the field of development: for instance: What sort of energy is produced and how is nuclear energy used versus new sources? How many companies can generate and distribute energy? The effects on the working structure must be evaluated. The studies about the need (and the opportunity) for human resources with new technical and professional abilities, that is, through the widening of competences and labour skills in a new context of the development of the social and human capital.

(8) More theories about international relations and regional cooperation-integration must be generated. It is true that the nation-state is pulled apart by different demands and requirements, but it is also true that the State in the IS will have to fulfil new functions and generate widened capacities in a scenario of growing complexity. In a word, there are two issues combined here: (i)

the IS, risks, threats and vulnerabilities, critical infrastructures, implementation of mechanisms and governance policies all under a sustainable development focus, and (ii) the international dimension, as they will be, over the next few years issues of great interest for the majority of countries and will be included in public and private agendas.

9.0 CONCLUSION

The central idea is that we have entered an era of complexity where instability, uncertainty and turbulence are not merely exogenous and circumstantial variables but also structural features of the system. This complexity cannot be studied in an abstract way when studying its harmful effects and, in particular, the risks induced by it.

The international context also presents deep asymmetries which considerably influence the intensity of the risks and the means used to counter them. This situation can be measured not only by the unequal levels of relative development or by the high concentration of the fruits of technical progress in developed countries, but also by the diverse capacities societies may have to respond to this complexity, as has been proven by some CEPAL-United Nations studies. It is clear from what has been exposed here, that the risks of infrastructures must be analyzed bearing in mind the social, political and economic characteristics specific to each community being studied.

In the Organization of Americans States (OAS) the countries of Latin America and North America agree a common vision of the risks and threats in the cyberspace and the firsts lines to protect and reaction the Critical Infrastructures physical and informational (virtual) against it. In Latin America this situation requires as an answer the elaboration of national, bilateral or multilateral policies and strategies to create strong capabilities. In the Region it is necessary firstly to secure Critical Information Infrastructures.

One of the fundamental issues in contemporary society is related to the matter of how to take decisions in matters concerning the whole of the social system. And, that these decisions might carry along negative aspects that could affect diverse social groups in different modes and degrees be it geographical, economic, educational and age aspects, etc.

It must also be considered that problems are multidimensional: there is a diversity of aspects and factors such as legal, economic, social, political, financial, health, psychological, environmental, technical… and then there are risks that accumulate with the passing of time. This is why there should always be options: there is no single way forward. Issues are difficult to solve and many questions require a deep societal and governmental awareness: who is responsible, who pays, who criticize, who takes decisions and how and above all, how the scientific base is used.

“Uncertainty is inherent in all stages of risk assessment”. Truly, according to what is maintained in recent studies about decision making, it is a problem of processes and institutions, of efficiency but also of precaution, trying to understand what are the potentialities and the limitations of the method and the evidence being employed. Both, processes and institutions must be designed for the objectives of risk analysis in a given situation.

Finally, it must be emphasised that one of the tasks to be developed is that of the rational thinking that is scientific, including knowledge, instruments and technologies, which will allow us to analyze, interpret, identify and manage risks. This may be the most important competitive advantage in the 21st Century together with financial and engineering capacities (amongst others) for the design, development and direction of these infrastructures.

REFERENCES

Aven, T. and O. Renn (2010). *Risk Management and Governance. Concepts, Guidelines and Applications*, Berlin, Springer Verlag.

BBC (2007). Estonia hit by 'Moscow cyber war'. BBS News. One-Minute World News, 17 May 2007.

Beck, U. (1992). *Risk Society. Towards a New Modernity*. London, Sage.

Bernal-Meza, R. y G. Masera (2007). “Sociedad de la información: etapa posterior de la globalización”, *Realidad Económica*, Nro. 227, pp. 90-116.

Bevir, M. (edit.) (2007). *Encyclopedia of Governance*. London, Sage.

Bischoff, H.J. (2008). *Risks in Modern Society*. Dordrech, Springer.

Campbell, S. (2007). ‘Risk and the Subjectivity of Preference’. *Journal of Risk Research,* Volume 9, 2006 - Issue 3, p. 225-242.

Castells, Manuel (2002). *‘Net war,’ Diario El País, Madrid*, Spain, 6/02/02.

Castells, Manuel (1998). *‘Net State?’: Economic and political Globalization in the information age’,* Society and State Reform, San Pablo, march, 1998.

Cepal (2003). *Los Caminos Hacia una Sociedad de la Información en América Latina y el Caribe*, Santiago de Chile, UN-CEPAL, Documento de Bávaro.

Comisión Europea (2003). *Hacia la Europa basada en el conocimiento. La Unión Europea y la sociedad de la información*. Luxemburgo: Oficina de Publicaciones Oficiales de las Comunidades Europeas.

Gheorghe, A. et. al. (2005). *Critical Infraestructures at Risk*, Dordretch, Springer.

Gheorghe, A.,M. Masera, and F.Polinpapilinho (edits.) (2013). *Infranomics. Sustainability, Engineering Design and Governance*, Dordrech, Springer.

Girard, B. and F. Perini (eds.) (2013). *Enabling Openness: The future of the information society in Latin America and the Caribbean*. Otawa: Canada, IDRC.

Hadkiewicz, W. and P. Gawowicz (2013). “Information technologies in the postindustrial society”. *Procedia - Social and Behavioral Sciences*, Elsevier, N°103: p. 500 – 505.

International Telecommunications Union [ITU], (2006). *WSIS Golden Book*, Geneva, United Nations.

International Risk Governance Council [IRGC] (2006), *White Paper on Managing and Reducing Social Vulnerabilities from Coupled Critical Infrastructures*, Geneva.

International Risk Governance Council [IRGC] (2009). *Risk Governance Deficits*
An analysis and illustration of the most common deficits in risk governance. Geneva.

International Risk Governance Council [IRGC] (2010). *Emerging risks. Sources, drivers and governance issues*, Geneva, Revised edition.

International Risk Governance Council [IRGC] (2016). *Workshop Report. Cyber Security Risk Governance.* Geneva. Swiss Re Centre for Global Dialogue, Zurich, Switzerland 29 – 30 October 2015.

Keyworth, G. (1998). *The Digital State: How State Governments are Using Digital Technology*, Washington, DC. The Progress and Freedom Foundation, September 1998. .

Lechte, J. (2003). *Key Contemporary Concepts*. London, Sage.

Lippert, T. H. (2016). NATO, Climate Change, and International Security: A Risk Governance Approach. Santa Monica, CA: RAND Corporation, 2016.

Mansell, R. (edit.) (2009). *The Information Society. Critical concepts in sociology*. London and New York, Routledge.

Masera, G. (2008). “Impactos en la Sociedad Global de la Información”, *Políticas Públicas*, FAE-Universidad de Santiago de Chile, vol.2, nro. 1, pp. 5-27.

Masera, G. (2010). *Epistemología y Economía Mundial*, Mendoza, EdUDA.

Mythen, G. (2004). Ulrich Beck. *A Critical Introduction to the Risk Society*. London, Pluto Press.

NIPP (2009). *National Infraestructure Protection Plan*. US Department of Homeland Security.Washington D.C., U.S. Department of Home Security.

Nye, J. and J. Donahue (edits.) (2000). *Governance in a Globalizing World,* Washington, D.C., Brookings Institution.

OAS (2003). Declaration on Security in the Americas, Mexico City, 28 October 2003.

OAS (2004). *Adoption* of a *Comprehensive*. *Inter-American Strategy* to *Combat Threats* to *Cybersecurity*: *A Multidisciplinary Approach to* Creating a Culture of *Cybersecurity*. Washington DC.

OAS (2007). *Declaration on the Protection of Critical Insfrastcture in the Hemisphere in the Face of Terrorism*. Inter-American Committee against Terrorism (CICTE), Panama City.

OAS (2014). *Cybersecurity trends in Latin America and the Caribbean,* Washington DC.

OECD (2003). *Emerging systemic risks in the 21st Century – An Agenda for Action*, Paris; International Futures Project.

Ortiz, J. (2008). *«Argentina the Challenge of the information operations». The World Views IO*. IO Sphere. San Antonio TX, Special Editon 2008, p. 57-63.

Ortiz, J. (2012). "Estrategia de Defensa Cibernética en la era de la información", *Revista de la Escuela Superior de Guerra del Ejército Argentino*, N° 582, (Septiembre-Diciembre), p. 89-112.

Peres, W. y M. Hilbert (eds.) (2009). *La sociedad de la información en América Latina y el Caribe*. Santiago de Chile, Naciones Unidas-CEPAL.

Pierre, O. and G. Peters (2005). *Governing Complex Societies. Trajectories and Scenarios*. New York, Palgrave Macmillan.

Renn, O. (2008). *Risk Governance. Coping with Uncertainty in a Complex World.* London, Esarthscan.

Saadawi. T. and L. Jordan (edit.) (2011). *Cyber Infrastructure Protection*. PA: USA, Strategic Studies Institute.

Thomas, Timothy L. (2005). *Ciber Silhouettes: Shadow Over Information Operations*, Foreign Military Studies Office, KS.

Thompson, P. (1990): "Risk Objectivism and Risk Subjectivism: When Are Risks Real?", *Risk: Health, Safety & Environment*, n°1: 3, (January), Article 4, p. 3-22.

Webster, F. (2002). "The information Society Revisited", in L. Lievrouw and S. Livingstone (Edit.). *Handbook of New Media, Social Shaping and Consequences of ICT's.* London, Sage, 2002, p. 22-33.

Webster, F. (2006). *Theories of Information Society*. London and New York, Routledge, third edition.

Reliability and Validity of Survey Scales

Louangrath, P. I. ★

About the Author
★Louangrath, P.I. is an Assistant Professor in Business Administration at Bangkok University, Bangkok, Thailand. He could be reached by email at: Lecturepedia@gmail.com

ABSTRACT
In this paper, we answered two questions: What is the reliability of a response scale in a question? What is the validity of a response scale in a question? The purpose of this paper is to present practical tools for measuring the reliability and validity of response scales used in written survey. Reliability measures consistency and validity measures precision. Our objective is to determine the reliability and validity of Likert and non-Likert scales used in research instrument. The data came from the numerical values of each type of scale. The Likert-type of scales include (1,2,3,4,5), (1,2,3,4,5,6,7) and (1,2,3,4,5,6,7,8,9,10). Non-Likert scale was (0,1,2,3). Reliability was measured by the estimated of λ under system analysis. The response space was proxied as a system to create a range between maximum and minimum values in the scale. Validity was tested by using the Fisher transformation of the estimated Z score of λ series. Empirical evidence shows that non-Likert scale (0,1,2,3) is 92% reliable while the Likert-type of scale had 90, 89, and 88% reliability. Validity test showed that non-Likert scale was 93% reliable, while the Likert-type scale had 89, 61, and 57% precision. Through Monte Carlo simulation and NK landscape method for optimization, the ability of information retention for non-Likert scale was 0.96 and 0.73, 0.75, and 0.77 for Likert scales. We standardize the scale efficacy in a 5.0 system, the non-Likert scale is 4.73 and 2.35, 2.45, and 2.41 for Likert scales.

Keywords: Likert, questionnaire, reliability, scale, survey, validity

JEL Code: C12, C13, C15, C18, C83, C93

CITATION:
Louangrath, P. (2018). "Reliability and Validity of Survey Scales." *Inter. J. Res. Methodol. Soc. Sci.*, Vol., 4, No. 1: pp. 50-62. (Jan. – Mar. 2018); ISSN: 2415-0371.

1.0 INTRODUCTION

Scientific research must meet two requisites: (i) reliability and (ii) validity. Reliability is defined as consistency in results from repeated measurements (Glasser *et al.*, 1990). Validity is defined as the precision of the result; precision is defined as minimal difference between the observed and expected value (BS 5497-1, 1979; and BS ISO 5725-1, 1994). In social science, these two requirements points to the instrument for data collection as the unit of analysis.

In social science, data are generally collected through the use of written questionnaires or survey. In order for the research findings to achieve empirical and scientific standing, the instrument must be properly calibrated. Instrument calibration means that it must pass reliability and validity tests. These tests answer two fundamental questions: (a) Is the instrument reliable by producing consistent results with repeated measurements? and (b) Is the instrument valid by achieving the precision in its measurement?

There are four possible combinations of outcome in reliability and validity in a given instrument testing. These four possible combinations of instrument reliability and validity may be explained by the Pearson 2x2 table. Firstly, an instrument may be reliable, but not precise. In this scenario, the instrument may produce achieve homogeneity in results when subjected to repeated measurements, but failed to in achieving precision (AD). The failure in precision may be evidence by significant difference between the observed value and the expected value. In a second scenario, the instrument may fail to produce homogenous results and also fails in precision (BD). In this case, the instrument is a complete failure. A third scenario may come from a possible outcome where the instrument does produce homogenous results, but passes precision testing (CB). This is possible if the respondents consistently misunderstand the question and consistently give wrong answers. This type of instrument is also faulty. A four scenario, where the instrument produces homogenous results and precise measuring, is an ideal instrument (AC).

Table 1: Possible outcome of reliability and validity test of an instrument

	YES	NO
Reliability	*A*	*B*
Validity	*C*	*D*

The test statistic for an ideal instrument (AC) may be accomplished by:

$$\chi^2_{df=1,95\%} = \frac{(n-1)(\text{ad}-\text{bc})^2}{(a+b)(a+c)(b+d)(c+d)} \quad (1)$$

where a, b, c, d are the frequency of items that falls into the YES and NO categories, and n is the number of response in a question in case of per question testing and number of questions in a survey in case of survey testing. The theoretical value is $\chi^2_{df=1} = 3.80$ for 0.95 confidence interval (Kanji, 2006, p. 85; and Yates, 1934).

2.0 LITERATURE REVIEW

2.1 Test for reliability

Reliability is defined as consistency in results after repeated measurements (Taylor, 1999). This consistency is defined by the homogeneity of the results. The argument in homogeneity is that among the k samples, their variances are equal. The decision rule is given by: $H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$ and $H_A : \sigma_i^2 \neq \sigma_2^2$. The alternative argument is that at least one of the paired variance is not equal. The existence of at least one unequal pair is a proof that the *k* samples lack homogeneity. There are two tests available for testing the homogeneity of data: (i) Lavene test and (ii) Barlett's Test.

2.1.1 Lavene Test for homogeneity

The first test for homogeneity is the Lavene's test. The Lavene test is given by:

$$W = \frac{N-k}{k-1}\left(\frac{\sum_{i=1}^{k} N_i\left(\bar{Z}_{i\bullet} - \bar{Z}_{\bullet\bullet}\right)^2}{\sum_{i=1}^{k}\sum_{j=1}^{N_i}\left(N_{ij} - \bar{Z}_i\right)^2}\right) \tag{1}$$

where Z_{ij} can be one of the following:

$Z_{ij} = \left|Y_{ij} - \bar{Y}_i\right|$ where $\bar{Y}$ is the mean of the i^{th} group,

$Z_{ij} = \left|Y_{ij} - \bar{Y}_{i_{med}}\right|$ where $\bar{Y}_{i_{med}}$ is the median of the i^{th} group, or

$Z_{ij} = \left|Y_{ij} - \bar{Y}_i'\right|$ where $\bar{Y}'$ is the 10% trim of the mean.

The decision rule for the test statistic is based of the critical value of the F table where homogeneity does not exists if $W > F_{\alpha,k-1,\mathrm{N}-1}$ at error level α and degrees of freedom $k-1$ and $N-1$.

2.1.2 Barlett Test for homogeneity

The basis for the argument under the Bartlett's test is the same as in Lavene: $H_0: \sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$ and $H_A: \sigma_i^2 \neq \sigma_2^2$. The test statistic of the Bartleet's test is given by (Snedecor and Cochran, 1989):

$$T = \frac{(N-k)\ln s_p^2 - \left(\sum_{i=1}^{k}(N_i - 1)\ln s_i^2\right)}{1 + \left(\frac{1}{3(k-1)}\right)\left(\sum_{i=1}^{k}\frac{1}{N_i - 1}\right) - \left(\frac{1}{N-k}\right)} \tag{2}$$

where s_i^2 = variance of the group; N = total sample size; N_i = sample size of the i^{th} group; k = number of groups; and s_p^2 = pooled variance. The pooled variance is obtained through:

$$s_p^2 = \sum_{i=1}^{k}\left(\frac{(N_i - 1)s_i^2}{N-k}\right) \tag{3}$$

The argument is that there is unequal variance if $T > \chi^2_{1-\alpha,k-1}$ where $k-1$ is the degree of freedom. Another expression of the Bartlett's test is written as:

$$M = v\ln S^2 - \left(\sum_{i=1}^{k} v_i \ln S_i^2\right) \tag{4}$$

where $v = \sum_{i=1}^{k} v_i$, $v_i = n_i - 1$, $S^2 = \sum_{i=1}^{k} v_i \div v$, $S_i^2 = \sum_{i=1}^{n} \left(X_{ij} - \bar{X}_{i\bullet} \right)^2 \div (n_i - 1)$. In this case M is distributed approximately χ_{k-1}^2 with a minimum sample size of $n < 5$. However, this M value is a biased estimated which requires a correction C where:

$$C = 1 + \frac{1}{3(k+1)} \left(\sum_{i=1}^{k} \frac{1}{v_i} - \frac{1}{v} \right) \tag{5}$$

Therefore, in evaluation homogeneity of a data set, instead of using M , the corrected Bartlett's test is simply: M / C. Thus, the hypothesis testing becomes: $H_0 : \frac{M}{C} \le \chi_{k-1}^2$ and $H_A : \frac{M}{C} > \chi_{k-1}^2$ (Dixon and Massey, 1969).

It is a common practice in social science to report Cronbach's alpha as the indicator for the reliability of the survey. The Cronbach's alpha is given by:

$$\alpha = \left(\frac{n}{n-1} \right) \left(1 - \frac{\sum v_i}{v_T} \right) \tag{6}$$

where n = number of items, v_i = variance of the score for the individual question, and v_T = variance of the entire score in the survey. (Cronbach, 1951). The function of the Cronbach's alpha is to verify the consistency of answers among respondents. It does not verify whether the instrument itself is reliable. The Cronbach alpha measures the consistency of responses, not the reliability of the instrument. The consistency of responses in a sample tells the similarity or closeness of response values in a group where the "group" is defined as the sample. A defective instrument may produce consistent responses among respondents and, thus, giving high Cronbach's alpha. In such a case, the reading of Cronbach's alpha is misleading. Although consistent, such responses may still be unreliable due to the defect of the instrument.

Cronbach's alpha has been misused by researchers who synomously confused internal consistency with homogeneity (Cortina, 1993; Nunally and Bernstein, 1994; Schmitt, 1996; Sitjsma, 2009; and Streiner, 2003). Internal consistency is not sufficient condition for homogeneity (Green, *et al.*, 1977). High alpha value may mean low validity due to redundancy of questions asking for similar responses at the expense of construct coverage (Boyle, 1985, 1991; and Kline, 1979). It is clear that alpha value is not indicator for reliability. As for its purported use for reliability, the standard reading has been that for basic research, the value should be $0.70 < \alpha < 0.80$ and for issues requiring high standard, the range should be $0.90 < \alpha < 0.95$. Nevertheless, these recommendations and use of Cronbach's alpha evaluated the entire survey does not assess the reliability of each question. This inadequacy was best expressed by Cronbach himself who wrote that:

> "I no longer regard the alpha formula as the most appropriate way to examine most data. Over the years, my associate and I developed the complex generalizability (G) theory." (Cronbach *et al.* (1963); Cronbach *et. al.* (1973); *see also* Brennan (2001); Shavelson and Webb (1991), which can be simplified to deal specifically with a simple two way matrix and produce coefficient alpha (Cronbach (2004), p. 403). *Cited in* N.M. Webb, R.J. Shavelson and E.H. Haertel (2006). "Reliability Coefficients and Generalizibility Theory." Handbook of Statistics, Vol. 26, p. 2. ISSN 0169-7161.

The testing of the scale used in the survey is a different issue than what Cronbach's alpha intended to serve. Cronbach's alpha tests the consistency of responses which are external to the instrument. The reliability or consistency of responses is not the same as the reliability of the instrument. Cronbach's alpha is not a tool for instrument calibration. This paper introduces the tool for testing the reliability and validity of the instrument by examining the response space in the individual survey question.

Reliability is defined as consistency in results. Validity is the precision of the measurement. We argue that if the instrument is reliable, the response would also be reliable even if not consistent. Respondents may give different responses to the same question and give rise to low Cronbach's alpha. Such scenario does not vitiate the reliability of the instrument. In case of validity testing, the precision of the instrument becomes the unit of analysis. Within a response space, there is an expected value that could be used as a threshold value to gauge the scale's precision. A survey scale whose observed value differs significantly from the expected value is an imprecise instrument. Such an instrument fails validity test. This paper presents two research questions: (i) what is the reliability of the survey scale? and (ii) what is the validity of the survey scale?

2.2 Test for validity

As a test for precision, the test for validity requires a threshold value against which the observed value from the survey scale may be assessed. We depend on the normal distribution curve as the basis for evaluating validity. Using the percentage probability, the mid-point of the percentage probability of 50% or a critical value of $Z = 0$ is used as the threshold value with $\pm 5\%$ error. The test statistic proposed for validity test is given by:

$$\hat{V} = \frac{F(Z) - 0.50}{\sqrt{\frac{n}{12}}} \qquad (7)$$

where $F(Z)$ is the percentage probability of the observed value of the scale, and n is the number of choices contained in the scale. For instance, a scale in a form of (0,1,2,3) has $n = 4$ and (1,2,3,4,5) has $n = 5$. Formula (7) is based on the optimization equation in NK landscape simulation.

3.0 DATA CONFINED TO ELEMENTS OF THE SCALES

The data used in this paper comes directly from the content of each scale. No outside data is necessary. There are four scales selected for the study. These scales are categorized into two types: Likert and non-Likert. Likert scales include (1,2,3,4,5) (Likert, 1932), (1,2,3,4,5,6,7) and (1,2,3,4,5,6,7,8,9,10). Non-Likert scale is (0,1,2,3).

Despite disagreement in the literature over the question of whther Likert scale is quantitative or qualitative (Jamieson, 2004; and Norman, 2010), we treat Likert and non-Likert as quantitative data. Likert scales do not contain zero. Thus, they could only be subjected to continuous distribution testing (Abramowitz and Stegun, 1972). Non-Likert scale, on the other hand, contains zero.

Discrete and continuous distribution may be used for testing non-Likert scale or scale containing zero value. The ability of the data set to allow the type of probability analysis is not a small issue for purposes of hypothesis testing. A data set that allows both discrete and continuous distribution testing affords researchers the flexibility and varied tools for hypothesis testing. The information obtained from such data set is more extensive compared to the inflexible Likert data set allowing only continuous distribution testing. The distinguishing characteristics of the Likert and non-Likert scales come from two factors: (i) the presence of zero in non-Likert scale, and (ii) the number of choices or the size of the subspace in each scale. These two characteristics are relevant in evaluating the efficacy of the scale.

Table 1: Common scales used in survey questionnaires

Scale Type	Scale Description	Data Distribution	Data Flexibility
Type 1: Non-Likert	(0,1,2,3)	Discrete & continuous	Yes
Type 2: Likert	(1,2,3,4,5)	Continuous	No
Type 3: Likert	(1,2,3,4,5,6,7)	Continuous	No
Type 4: Likert	(1,2,3,4,5,6,7,8,9,10)	Continuous	No

4.0 METHODOLOGY

4.1 Linearize the scale by QQ plot

The scale data is subjected to QQ plotting in order to obtain a linear equation: $Y = a + bX$. The linear equation will be used to obtain the expected value for each element in the cale. The QQ plot starts with calculating the time function:

$$F(t) = \frac{i - 0.30}{n + 0.40} \tag{8}$$

where i = item of sequential listing of the scale elements, and n = number of items in the scale. For example, (0,1,2,3) has $n = 4$, (1,2,3,4,5) has $n = 5$, (1,2,3,4,5,6,7) has $n = 7$, and (1,2,3,4,5,6,7,8,9,10) has $n = 10$. Since these items occur in sequence, the time function $F(t)$ is used as a starting point to obtain the predictive model for the scale. Dependent (Y) and independent (X) variables in the QQ plot, to create the linear equation, are obtained by:

$$X_{qq} = \ln\left(\ln\left(\frac{1}{1 - F(t)} \right) \right) \tag{9}$$

$$Y_{qq} = \ln(X_{scale}) \tag{10}$$

In order to construct a linear equation in a form of $Y = a + bX$, the following basic statements are required:

$$\begin{aligned} I &= n\sum XY - \sum X \sum Y \\ II &= n\sum X^2 - \left(\sum X\right)^2 \end{aligned} \tag{11}$$

The slope of the line is obtained by $b = I \div II$ and the intercept is obtained by $a = \overline{Y} - b\overline{X}$. With known linear equation, the expected value for each element of the scale subspace is estimated to be $\hat{Y}_j$.

With known expected value series of $\hat{Y}_j$, the CDF and PDF of each element of the scale is determined. The CDF is given by the Z score equation:

$$Z = \frac{\hat{Y}_j - \dot{Y}}{S_{\hat{y}}} \tag{12}$$

where $\hat{Y}_j$ = each expected value in the *j* group or response subspace, $\dot{Y}$ = mean value of the *j* group, and $S_{\hat{y}_j}$ = standard deviation of the *j* group. The Z score represents the critical value for which the percentage probability may be read from the Z table. The corresponding percentage probability from the Z table is called cumulative distribution probability (CDF or $\phi(z)$). This series of percentage probability is used to obtained *lambda* or the reliability of the scale.

4.2 Reliability testing of scales used in survey

The reliability of the scale is obtained through the calculation of lambda. In order to obtained lambda, it is necessary to know the value of CDF and PDF or probability distribution function. The PDF may be obtained by:

$$PDF = \frac{\Phi(z_2) - \Phi(z_1)}{\hat{y}_2 - \hat{y}_1} \quad (13)$$

Now, lambda may be determined by:

$$\lambda = \frac{CDF}{1 - PDF} \quad (14)$$

The value of λ will come in series of *k* elements in the scale. For instance, if the scale is (0,1,2,3), *k* element is equal to $k = 4$. The reliability of the scale is equated to the stability of the λ which is estimated to be: $\overline{\lambda} + s_\lambda$ or the mean of λ plus its standard deviation.

While $\overline{\lambda} + s_\lambda$ represents the reliability of the scale with *k* elements, the value of R may be converted to critical value in the Z table in order to obtain the level of confidence in the reliability of $\overline{\lambda} + s_\lambda$. Thus, with known R in the range of $0 < R < 1$, trace the value in the Z table for the corresponding critical value. The critical value of Z is subjected to Fisher transformation in order to calculate the upper and lower limits of the error for the estimation. The Fisher transform is obtained through:

$$Z = 0.50 \ln\left(\frac{1+R}{1-R}\right) \quad (15)$$

This transformed Z is an estimate; thus, it must carry a range of error. The upper and lower error limits of the transformed Z value is obtained through:

$$\xi_u = Z + \hat{Z}\sqrt{\frac{1}{n+3}} \quad (16)$$

$$\xi_l = Z - \hat{Z}\sqrt{\frac{1}{n+3}} \quad (17)$$

where Z = observed value and $\hat{Z}$ = threshold value at a specified percentage confidence interval. The range of the reliability R of the scale becomes $R + \xi_u$ for the upper value and $R - \xi_u$ for the lower value. This range will later provide the maxima and minima for Monte Carlo simulation.

4.3 Validity testing of scales used in survey

In order to determine the validity of the scale, the Fisher transformed Z and the critical Z value read from the table, corresponding to R, is used. This level of confidence of R is obtained in two steps, firstly:

$$Z^* = \frac{Z - \bar{Z}}{S} \quad (18)$$

where Z^* = critical value read from the Z table corresponding to R, $\bar{Z}$ = Fisher transformed Z in equation (9), and $S = \sqrt{1/(n-3)}$.

The value of Z^* is in a form of critical value which is compared to the theoretical value of $Z_\theta = 0.50$ or the mid-point of the distribution curve. The critical point at 0.50 is the mid-point of the distribution curve; we use this point as the threshold value for precision measurement. The difference $Z^* - Z_\theta$ represents the error or pValue of the estimated validity and $1-(Z^* - Z_\theta)$ represents the validity of the scale. If scale is precise, the value of $Z^* - Z_\theta$ must be zero or close to zero within an acceptable level of error, i.e. located "bulls eye" at the center of the distribution curve. Thus, $\Pr[Z^* - Z_\theta] \le \Phi(z)$ where $\Phi(z) \le 0.05$.

4.4 Monte Carlo simulation and NK landscape optimization

In order to increase the accuracy of the estimated results of reliability and validity tests for the survey scales, Monte Carlo simulation was used to obtain an optimum score. The optimum score was determined by NK landscape optimization method. Monte Carlo simulation began with the determination of the number of repetitions for the measurement, thus:

$$N = \left(\frac{3\sigma}{E}\right)^2 \quad (19)$$

where σ = the estimated standard deviation of the three components of Monte Carlo numbers: $x_1 = \max$, $x_1 = \min$, and $x_1 = (\max - \min)/2$, and $E = ((\max - \min)/2) \div 50$. The value series came from the *k* group or $\lambda_1, ..., \lambda_k$ in reliability and $\hat{Z}_1, ..., \hat{Z}_k$ for validity precision testing.

Since the Monte Carlo simulation provides a good estimate, the estimated values of reliability and validity are then used to find the optimum points. Optimization was accomplished by using NK landscape method:

$$Z_{opt} = \frac{F(\mathrm{x}) - 0.50}{\sqrt{\frac{n}{12}}} \quad (20)$$

where $F(x) = 1 - error$ and $error = U - 0.50$. The value for U is described in (22) and n = number of answer choice in each survey question.

The observed optimum value is compared to the theoretical value. The theoretical value is obtained by:

$$Z^* = \frac{(\mu + \sigma) - 0.50}{\sqrt{\frac{N}{12}}} \quad (21)$$

where μ = expected mean within the scale per survey question, and σ = expected standard deviation within the scale.

4.5 Prospect theory for validity level

Since each scale consists of quantitative elements in sequence: (0,1,2,3), (1,2,3,4,5), (1,2,3,4,5,6,7) and (1,2,3,4,5,6,7,8,9,10), the weight and probability of each may be calculated. The weight is determined by $1/n$ in each scale. Each response choice is accorded with equal weight. The probability is given by the CDF of each element of the scale. With known weight, probability and value of each element in the scale, the prospect of each value may be calculated. The Kahnman-Tvertsky prospect theory is used for this purpose:

$$U = \sum w_i p_i x_i \quad (22)$$

where w = weight of each element in the scale; p = percentage probability of each element; and x = observed value of the scale element.

The prospect U serves as the basis for validity testing. Recall that validity is the test for precision or level of accuracy (Metz, 1978). The threshold for precision in a distribution is located at the center of the distribution curve or $Z = 0.50$. The U value must be contrasted with the mid-point of the distribution curve in order to verify the precision level of the survey scale, i.e. $U - 0.50 = error$. This $error$ is then used to obtain the confidence level $1 - error$. The confidence level is the precision level. This precision level defined validity.

Lastly, we ask another question: how much does the scale retain information? What is the probable information retention rate? Conversely, what is the probable information loss for each type of scale? The answers are obtained through the use of NK landscape optimization method (14). This last information was used to score the survey scale. Each survey scale was subjected to a 5-points scoring system of precision where 5 = highest, 4 = high, 3 = medium, 2 = low and 1 = very low. The rational of this scoring system is to convert the validity level into a 5-points nominal data for easy understanding.

5.0 FINDINGS AND DISCUSSION

The findings are presented in two parts. Part 1 is the raw estimate of the reliability (5.1) and validity (5.2) for the scales. Part 2 is the theoretical values of reliability and validity obtained through Monte Carlo simulation and NK landscape optimization method (5.4). In section 5.5, we carried the discussion from a single question testing to the entire instrument (global environment) based on the findings in per question subspace (local environment).

5.1 Non-Likert scale is more reliable

Non-Likert scale (0,1,2,3) is more reliable because its λ series is more stable. This stability indicates consistency of the system. The system is defined by elements of the response sequence in the scale. Since λ is an estimate, it is subjected to error. Thus, we use $\lambda + s$ as the indicator for reliability. Under this method, non-Likert scale (0,1,2,3) scored highest followed by (1,2,3,4,5), (1,2,3,4,5,6,7) and (1,2,3,4,5,6,7,8,9,10), respectively.

Table 2. Reliability of various scales

Item k elements	Scale 1 Non-Likert Type (0,1,2,3)	Scale 2 Likert Type (1,2,3,4,5)	Scale 3 Likert Type (1,2,3,4,5,6,7)	Scale4 Likert Type (1,2,3,4,5,6,7,8,9,10)
1	0	1	1	1
2	1	2	2	2
3	2	3	3	3
4	3	4	4	4

5	-	5	5	5
6	-	-	6	6
7	-	-	7	7
8	-	-	-	8
9	-	-	-	9
10	-	-	-	10
R	0.92	0.90	0.89	0.88
$1-R$	0.08	0.10	0.11	0.12
pValue	0.07	0.11	0.39	0.43
Conclude	Pass	Fail	Fail	Fail

5.2 Validity of survey scale

The pValue reported in Table 2 measures the error level. The error is the indication of how far $Z*$ is located from the center of the distribution curve or $Z_\theta = 0$; this is the midpoint of the symmetrical normal curve. We found that non-Likert scale is 0.07 points further from the center while other Likert-type scales were 0.11, 0.39 and 0.43 further away from the center. Since $Z_\theta = 0$ is the threshold for precision, the closer the critical value of the scale's reliability score is to $Z_\theta = 0$, the more precise or valid is the scale. In this case, a non-Likert scale in a form of (0,1,2,3) is more precise than all three forms of Likert scales: (1,2,3,4,5), (1,2,3,4,5,6,7) and (1,2,3,4,5,6,7,8,9,10).

5.3 Simulation confirm reliability and validity tests

The reliability series consists of the reliability indicator for the four types of scale: 0.92, 0.90, 0.89, and 0.88. The validity series consists of the confidence level determined by $1-error$ in Table 4. Each series was subjected to Monte Carlo simulation in order to obtain the expected value. The purpose is to verify the significance of the Z values under (14) and (15) with the following decision rule: $H_0 : Z_{opt} \le Z*$ and $H_0 : Z_{opt} > Z*$

The results of the simulations for reliability and validity show that the observed optimized values were larger than the estimated. In both cases, the difference $\left(Z_{opt} - Z*\right)$ is equal to 0.01. For reliability test simulation, the Monte Carlo repetitions were 635.96 and 28.15 repetitions for validity test. The errors between the observed and estimated values were 0.01 in both cases. Since this error is less than 0.05, we found no significant difference between the simulated and observe values obtained through the method proposed in this paper.

Table 3: Optimum reliability and validity under simulation

Test Type	Series	Monte Carlo N	Z_{opt} Observed	$Z*$ Estimate	$\left(Z_{opt} - Z*\right)$
Reliability	λ_k	635.96	0.06	0.05	0.01
Validity	R_j	28.15	0.27	0.26	0.01

The validity test is based on the Kahnman-Tvertsky prospect U where each U is contrasted with 0.50 threshold value. The value 0.50 is used as the threshold because it represents the mid-point of the distribution curve. The difference between the observed U and 0.50 or $U - 0.50 = error$ represents the inaccuracy of the scale and $1-error$ represents the precision level or validity. The result of this calculation is reported in Table 4.

Table 4: Prospect of information retention and information loss by scale type

Survey Scale Type	Prospect U	Error $U-0.50$	Confidence $1-error$	OPT Informatics	Information Loss
Type 1: $n=4$*	0.54	0.04	0.96	0.79	0.21
Type 2: $n=5$	0.77	0.27	0.73	0.36	0.64
Type 3: $n=7$	0.75	0.25	0.75	0.33	0.67
Type 4: $n=10$	0.73	0.23	0.77	0.30	0.70

*Scale types: Type 1 = (0,1,2,3); Type 2 = (1,2,3,4,5); Type 3 = (1,2,3,4,5,6,7); and Type 4 = (1,2,3,4,5,6,7,8,9,10).

Under NK landscape optimization simulation method, scale Type 1 (0,1,2,3) score highest in the prospect of information retention and lowest in the prospect of information lost. This testing is outside of the scope of reliability and validity. Whereas reliability test optimum reliability and validity in Table 3 shows the intrinsic property of the scale, the optimum prospect simulation gives an indication for the amount of information obtained through the survey scale. This latter piece of information may be another indicator for scale quality evaluation. If reliability test tells us how much does the scale produce consistency in response and validity test provides the level of precision within the scale subspace, then the utility prospect measurement provides the amount of information obtained within the scale response subspace. The efficacy of the scale under the prospect U is further converted into a 5-points score system by:

$$SCALE_{score} = OPT\left(\ln\left(N\alpha^2\right)\right) \quad (23)$$

where $OPT = \left(F(Z)-0.50\right) \div \sqrt{n/12}$, N = Monte Carlo repetition counts, and α = significance level. The scale score is given in a range of $0 < SCALE_{score} < 5$ The results of the calculation show that scale (0,1,2,3) has the highest scale score of 4.73 followed by (1,2,3,4,5,6,7) at 2.45, (1,2,3,4,5,6,7,8,9,10) at 2.41 and (1,2,3,4,5) at 2.35.

Table 5: Efficacy score for survey scale

Survey Scale Type	Monte Carlo N repetition	Log Monte Carlo $k=\ln\left(N\alpha^2\right)$	OPT Info. Retention	Scale score On 5-points System
Type 1: $n=4$	158,548.76	0.79	5.98	4.73*
Type 2: $n=5$	281,864.46	0.36	6.56	2.35
Type 3: $n=7$	634,195.04	0.33	7.37	2.45
Type 4: $n=10$	1,426,938.84	0.30	8.18	2.41

*The 5-points are defined as: 5 = highest; 4 = high; 3 = medium; 2 = low; and 1 = lowest.

5.4 Evaluating entire survey under Prospect Theory

Thus far, we have presented the evaluation of the scale in a survey question. With known per question reliability, validity and optimized prospect, we could estimate the same measurements for the entire survey with m number of questions. The estimate survey reliability may be given by the DeMoivre-Laplace theorem (Walker, 1985):

$$Z = \frac{X - np}{\sqrt{npq}} \quad (24)$$

where $n = m$ or number of questions in a survey, p = probability of success read from R for reliability and λ for validity, and $q = 1 - p$. The test number of questions are 10, 20, …, 100 in one survey. The answer obtained (24) answers the questions of "what is the maximum number of questions in a survey in order to maintain the level of R and λ at 0.95 confidence level?" The results of the test are presented in Table 6.

Table 6: Maximum number of questions in a survey

Scale Type	Reliability R	Max Questions without Distorting Reliability	Validity λ	Max Questions without Distorting Validity
(0,1,2,3)	0.92	20	0.96	30
(1,2,3,4,5)	0.90	20	0.73	10
(1,2,3,4,5,6,7)	0.89	40	0.75	20
(1,2,3,4,5,6,7,8,9,10)	0.88	50	0.77	30

Note that the answers provided in table 6 do not answer the question: "what is the reliability of the survey with *m* number of questions?" Instead, Table 6 answer the question of "what is the maximum number of question could a survey have in order to maintain 95% confidence levels of its reliability and validity?" The answer to this question is reserved for Part 2 of this paper. In Part 1 of the research, we limit the scope of the paper to per question assessment.

6.0 CONCLUSION

Non-Likert type of scale in a form of (0,1,2,3) has the highest level of reliability and validity. Likert-type scales: (1,2,3,4,5), (1,2,3,4,5,6,7) and (1,2,3,4,5,6,7,8,9,10), fail both reliability and validity tests. Reliability and validity are the cornerstones of scientific research. The reliability and validity testing tools presented in this paper are valuable for calibrating research instrument in social science. An unreliable instrument produces unreliable data. Unreliable data leads to faulty analysis and conclusion. Imprecise instrument could not bring the researcher close to the answer of the research question. Thus, it is important for researchers to use instruments that are reliable and valid. We reject Cronbach's alpha as inapplicable for instrument calibration. This paper provides practical tools for researchers to test the reliability and validity of the survey. Such tools are valuable contribution to research methodology in social science.

REFERENCES

Abramowitz, M. and Stegun, I. A. (Eds.) (1972). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, 9th printing. New York: Dover, pp. 927 and 930.

Boyle, G. J. (1985). "Self-report measures of depression: Some psychometric considerations." *The British Journal of Clinical Psychology*, **24**(1), 45-59. doi:10.1111/j.2044-8260.1985.tb01312.x

Boyle, G. J. (1991). "Does item homogeneity indicate internal consistency or item redundancy in psychometric scales?" *Personality and Individual Differences*, **12**(3), 291-294. doi:10.1016/0191-8869(91)90115-R

BS 5497-1 (1979). "Precision of test methods. Guide for the determination of repeatability and reproducibility for a standard test method."

BS ISO 5725-1 (1994). "Accuracy (trueness and precision) of measurement methods and results - Part 1: General principles and definitions." p.1.

Cortina, J. M. (1993). "What is coefficient alpha? An examination of theory and applications." *The Journal of Applied Psychology*, **78**(1), 98-104. doi:10.1037/0021-9010.78.1.98

Cronbach, L. J. (1951). "Coefficient alpha and the internal structure of tests." *Psychometrika*, **16**(3), 297-334. doi:10.1007/BF02310555

Dixon, W. J. and Massey, F.J. (1969). Introduction to Statistical Analysis, McGraw-Hill, New York.
https://archive.org/details/IntroductionToStatisticalAnalysis

Glasser, Mark; Mathews, Rob; Acken, John M. (June 1990). "1990 Workshop on Logic-Level Modelling for ASICS." *SIGDA Newsletter*, **20** (1).

Green, S. B., Lissitz, R. W., & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. Educational and Psychological Measurement, 37, 827-838. doi:10.1177/001316447703700403

Jamieson, Susan (2004). "Likert Scales: How to (Ab)use Them." *Medical Education*, **38**(12), pp.1217-1218.

Kanji, Gopal K. (2006). *100 Statistical Tests*; p. 85, 3rd ed., SAGE.

Kline, P. (1979). *Psychometrics and psychology*. London, United Kingdom: Academic Press.

Levene, H. (1960). In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, I. Olkin *et al.* eds., Stanford University Press, pp. 278-292.

Likert, Rensis (1932). "A Technique for the Measurement of Attitudes." *Archives of Psychology*. **140**: 1–55.

Metz, C.E. (October 1978). "Basic principles of ROC analysis." *Sem. in Nucl. Med.*, **8** (4): 283–98.

Norman, Geoff (2010). "Likert scales, levels of measurement and the "laws" of statistics". *Advances in Health Science Education*. **15**(5) pp. 625-632.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994) Psychometric theory (3rd ed.). New York, NY: McGraw-Hill, Inc.

Schmitt, N. (1996). "Uses and abuses of coefficient alpha." *Psychological Assessment*, **8**(4), 350-353. doi:10.1037/1040-3590.8.4.350

Sijtsma, K. (2009). "On the use, the misuse and the very limited usefulness of Cronabch's alpha." *Psychometrika*, **74**(1), 107-120. doi:10.1007/s11336-008-9101-0

Snedecor, George W. and Cochran, William G. (1989), Statistical Methods, Eighth Edition, Iowa State University Press.

Streiner, D. L. (2003). "Starting at the beginning: An introduction to coefficient alpha and internal consistency." *Journal of Personality Assessment*, **80**(1), 99-103. doi:10.1207/S15327752JPA8001_18

Taylor, John Robert (1999). An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. University Science Books. pp. 128–129. ISBN 0-935702-75-X.

Yates, F. (1934). "Contingency table involving small numbers and the $\chi 2$ test." *Supplement to the Journal of the Royal Statistical Society*, **1**(2): 217–235. JSTOR 2983604. https://www.jstor.org/stable/2983604

Walker, Helen M (1985). "De Moivre on the law of normal probability." *In* Smith, David Eugene. A source book in mathematics. Dover. p. 78. ISBN 0-486-64690-4. https://www.york.ac.uk/depts/maths/histstat/demoivre.pdf

Direct and Indirect Tests for Randomness

Chanoknath Sutanapong

About the author

★Chanoknath Sutanapong is an independent researcher. She may be reached by email at: chanoknath.sutanapong@gmail.com

ABSTRACT

The purpose of this paper is to explain the importance of randomness in data analysis. We point out the difference between random data and random selection. The scope of this paper is limited to randomness in discrete data. We explain the various tests used by NIST as the basis to understand randomness testing in discrete data. We also present an indirect test for randomness by relying on trend testing. Randomness is defined as lacking predictable pattern. Predictable pattern may be seen through trend. We argue that if there is a significant trend, the series is not random and vice versa. In this paper, we claim that trend test may be used an indirect test for randomness.

Keywords: Discrete data, randomness, sampling method, trend

CITATION:

Sutanapong, Chanoknath (2018). "Direct and Indirect Tests for Randomness." *Inter. J. Res. Methodol. Soc. Sci.*, Vol., 4, No. 1: pp. 63–83. (Jan. – Mar. 2018); ISSN: 2415-0371.

1.0 INTRODUCTION

1.1 Introduction to random data and random selection

Random selection deals with the sampling method. The misunderstanding here is that many researchers incorrectly believe that random selection produces random data. A sampling method does not produce the type of data nor can it change the nature of the data. If the data is random, no matter what kind of sampling method is used, the data would remain random. This randomness may be revealed through random testing. On the other hand, if the data is non-random, no matter what kind of sampling method is employed, the data would also remain non-random. For that reason, the use of "random sampling" method does not guarantee that the data would be random. The benefit of random sampling is to avoid selection bias. Selection bias occurs when each element of the population does not have equal chance of being selected. The statement "equal chance of being selection" coincidentally is a common definition of random selection; thus, to avoid the accusation of selection bias, researcher tends to claim to have used random selection.

In order to impose equal probability selection or random selection, it is necessary that the population size be known. This requirement is impractical because in most circumstances in social science, the population size is either unknown or if it is known, it is imprecise because population is dynamic, i.e. drop out, death, new birth and migration. Therefore, true random is not possible. The most that one can do is to "assume" that the population is non-dynamic, and the selection is made as if the population is finite. In such a case, the so-called random sampling is a mere presumption. The sampling made is no more than convenience sampling; it is only through self-fulfilling claim that it is said "randomly selected." Such a claim should not escape heighten ethical scrutiny. More attention should be given to randomness of the data, not how the data was selected.

Random data set means that the statistical test verifies that the data elements are random. The clam to randomness is substantiated through statistical test. The conclusion made is different from what had been said in the case of random selection. The claim by random test states that "the data came from a random process." It is this random process that is the holly grail in statistics because common statistical test requires that the data be random. "Be random" here means that the data comes from a random process, and random process means that the condition from which the data was sampled had been random. This is a completely different language from making a claim that the sampling method used was random selection. The focus on the randomness of the data under statistical tests does not give homage to the sampling method. As for selection bias, it would also show in the test as non-randomness in the data if there was indeed bias. Even if there is an alleged or suspicion of selection bias, if the data is indeed random, as verified by one or more of statistical tests, then such allegation and suspicion is vitiated.

1.2 Measurement of randomness

The objective of this section is concerned with the measurement of randomness. The methods used in the measurement of randomness are many. They can be categorized into data types, namely: (i) univariate, (ii) bivariate, and (iii) times series. In univariate data, the numbers come in one sequence or string. The test is imposed upon the string to verify whether the string or sequence is random; if so, it is said that the numbers come from a random process. Generally, in bivariate analysis, the sequence is dichotomize into (1, 0) where 1 is defined as category of interest or "success" and 0 is the category of non-interest of "failure." The dichotomy either comes from a true binary data in a form of (Yes | No) or it is made "dichotomous" through the application of mathematical expectation where the expected value or the mean is used as a separator dividing the sequence of numbers into two groups: below the mean and equal to and greater than the mean.

In the bivariate case, the data comes from two data array that may be categorized into array X and array Y. In such a case, the proof for randomness is verified through proving that X_i is comprised of independent elements $x_1, x_2, ... x_n$. Similarly, the same test is used to verify that the array Y_i also contain independent elements of $y_1, y_2, ... y_n$. Moreover, a third test in bivariate is to attest that X_i and Y_i are independent, i.e. the occurrence of Y_i does not depend on X_i so as to make the value of Y_i predictable with a given value of X_i.

A third for of number sequence used is randomness analysis is time series data. Time series are bivariate data consisting of X_i independent variable and Y_i dependent variable where the relationship between these two variable may be expressed by the regression model, thus: $X(t) = \beta_0 + \beta_1 X_1 + ... + B_n X_t + \varepsilon_i$.

In this model, it is no longer of interest as to the causality of Y_i, i.e. whatever explains the occurrence of Y_i is no longer the object of analysis. Whatever the causality or relation of X_i to Y_i is, it is taken for granted. Thus, whether the value of Y_i is produced, treat such series of events: $y_1, y_2, ..., y_n$ which occurs at time $t_1, t_2, ..., t_n$ respectively as $y_{t_1 = X(t_1)}, y_{t_2 = X(t_2)}, ..., y_{t_n = X(t_n)}$ and thus the event is listed simply as: $X(t_1), X(t_2), ..., X(t_n)$. This is called time series data array. In this scenario, the test for randomness is to look for the independents among the error term of each $X(t_i)$. If the series of $\varepsilon_i : \varepsilon_1, \varepsilon_2, ..., \varepsilon_n$ are statistically independent, it is said that the arrays of $X(t_i)$ are independent and, therefore, are random. Under these circumstances, autocorrelation test is used as a tool for verifying randomness among the time series $X(t_i)$.

2.0 Trend tests as indirect testing for randomness

If a common definition and characteristic of randomness is the lack of predictable trend or pattern, the most practical place to start the discussion of randomness is trend analysis.

One cautious note about trend test, the acceptance or rejection under the trend test is based on statistical significant test, i.e. the trend may exist but it is not considered statistically significant. Under this standard, the definition of randomness is relaxed because for cases where the trend test rejects, they may still manifest some kind of patterns, but that pattern does not pass the statistical significant test. This means that if the confidence interval of 0.95 is used, a pattern that would pass at 0.90 would be rejected at 0.95 as 'non-significant trend.' This does not necessary mean that pattern recognition of a data set of 0.90 trend display does not show a pattern. In fact, a pattern failing to reject a null hypothesis using 0.95 confidence interval may still be a "trend" or "pattern" but just does not pass 0.95 confidence interval test. For this reason, the various trend tests are not definitive test to define whether the data is random for failing to manifest trend pattern. Nevertheless, the introduction of trend test puts randomness in perspective to the extent that trend is defined as a recognizable pattern and that randomness is defined as the lack of recognizable pattern.

There are three trend tests commonly used in the field, namely (i) reverse Arrangement Test (RAT), (ii) Military Handbook Test (MHT), and (iii) La Place Trend test (LTT). The RAT method is given by:

$$Z = \frac{R - \left(\frac{r(r-1)}{4}\right) + 0.50}{\sqrt{\frac{(2r+5)(r-1)r}{72}}} \qquad (1)$$

where R = reversal counts, and r = repair time. The decision is governed by the following tables depending on whether we are looking for improving trend of degrading trend. The first type of trend is called an improving trend. The critical value of improving trend is given by the table below.

Table 1: One-sided test for improving trend

Repair: *r*	R(90%)	R(95%)	R(99%)
4	6	6	-
5	9	9	10
6	12	13	14
7	16	17	19
8	20	22	24
9	25	27	30
10	31	33	36
11	37	39	43
12	43	46	50

Source: http://www.itl.nist.gov/div898/handbook/apr/section2/apr234.htm

A second scenario of trend is a degrading trend where the pattern indicates that the process is deteriorating or "things are getting worse." The deteriorating trend uses the following critical values for testing its significance.

Table 1: One-sided test for degrading trend

Repair: *r*	R(90%)	R(95%)	R(99%)
4	0	0	-
5	1	1	0
6	3	2	1
7	5	4	2

8	8	6	4
9	11	9	6
10	14	12	9
11	18	16	12
12	23	20	16

Source: http://www.itl.nist.gov/div898/handbook/apr/section2/apr234.htm

The MHT approach is used when the values comes from a system that is described by Power Law. The Military handbook method is given by:

$$\chi^2_{2r} = 2\sum_{i=1}^{r} \ln\left(\frac{T_{end}}{T_i}\right) \quad (2)$$

where $T_i : T_1, T_2, ..., T_{end}$ are repair time counts. It is equivalent to *r* in the RAT equation above.

The third type of trend test is called the Laplace Trend Test. The LTT method is used when the data comes from an exponential system. LTT is given by:

$$Z = \frac{\sqrt{12r}\sum_{i=1}^{n}\left(T_i - \frac{T_{end}}{2}\right)}{rT_{end}} \quad (3)$$

If the data set shows that there is a significant trend under any of the three tests mentioned above, it is conclusive that the data is not random. The use of the trend test may be proach to verify if the data is random since "randomness" lacks predictable pattern.

3.0 NIST approach to randomness testing of discrete data

The national Institute of Standards and Technology (NIST) is an agency under the U.S. Department of Commerce (Ministry of Commerce). The mission of NIST is to "[p]romote U.S. innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life." (Perry, 1953; p. 123). In random testing, NIST offers 15 different tests to verify whether a sequence of number is random.

The 15 tests offered by NIST is based on binary digits or categorical data test based on (1, 0) discrete data. Although these 15 tests are considered sufficient for binary data, they are not adequate to deal with continuous data and times series data. For this reason, continuous and tome series data are treated outside of NIST's 15 tests. In this paper, we explain the first three of the fifteen tests in the NIST battery of test for randomness. The National Institute of Standards and Technology (NIST) offers a battery of 15 tests to verify randomness. These 15 tests are:

1. Frequency (monobit) test;
2. Frequency test within block;
3. Runs test;
4. Tests for the longest-run-of-ones in a block;
5. Binary matrix ranked test;
6. Discrete Fourier transform (spectral) test;
7. Non-overlapping template matching test;
8. Overlapping-template matching test;
9. Maurer's "university statistical" test;
10. Linear complexity test;
11. Serial test;
12. Approximate entropy test;

13. Cumulative sum test;
14. Random excursion test; and
15. Random excursion variant test.

In order to use the random verification method advocated by NIST, the data has to be discrete or categorical comprising of (1, 0). If the data comes from a continuous scalar form, such as a survey instrument answer choice of: (0, 1, 2, 3), it is still possible to make the continuous scalar into a discrete form by separating the zero and non-zero into two categorical that could then be mapped into the form (1, 0) required by NIST. For instance, (0, 1, 2, 3) could be re-categorized into discrete form by equating {1, 2, 3}, non-zero values, to 1 = 1, 2 = 1 and 3 = 1. The data set now becomes (1, 1, 1, 0).

In the alternative, other methods for random testing that does not fall under the 15 methods proposed by NIST could accommodate continuous data set without having to go through discretization process described above. For instance, the adjacent test deals with non-categorical data set on "as is" basis with the assumption that there is an assumed cloud within which the randomness is defined. If the test shows that the observed score falls outside of the cloud, it is considered non-random. The "cloud" is defined by an interval consisting of the lower and upper values. It is said that, if the observed value under the adjacent test falls below the lower limit, it is conclusive that the data comes from a non-random process. However, if the values fall outside of the "upper limit," it is not inconclusive in making the conclusion whether the data comes from a non-random process.

In both cases, NIST's approach and other approaches, it is important to note that the study of randomness focuses on the "data", i.e. the sample, not the method by which the data had been selected, i.e. "sampling method." This distinction, and the point where the emphasis is put when deciding on randomization, underscores the ineffectual insistence on random selection. Even if the data had been produced by random selection, if the data fails the random test that uses the data as the testing value, random selection could produce a data set that is non-random. The only answer that random selection could provide is that the selection process is non-biased. This rationale for random selection is still weak. Assume that the data is random; no matter how the data is selected the sample would still show that it is random data. That is, if by its nature the data is random, even if a convenience sampling method is used to select the data, it would still produce a random data. The nature of the data does not change as the result of the sampling method.

Numerical examples will be given to illustrate each test. In cases where the original data is not binary (1, 0) or (Yes | Not), the data string will be transformed into (1,0) using the expected value as the reference point, i.e. if $0 \le x_i < \bar{X}$ then 0 and $\bar{X} \le x_i$, or simply:

$$p(X) = \begin{cases} \bar{X} \le x_i \\ 0 \quad \text{Otherwise} \end{cases} \tag{4}$$

In the second case where bivariate X and Y are involved, autocorrelation and correlation coefficient tests will be used.

In the study of randomness, it is important to make a distinction between sampling method and data from the sample. The random test is the test of the data in order to answer the question of whether "the data came from a random process?" This is different from random sampling. Random sampling is a sampling method which is classified as *equal probability sampling*. It is given by:

$$P(sample) = \begin{cases} \binom{N}{n}^{-1} \text{ if } n(s) = n \\ 0 \quad \text{otherwise} \end{cases} \tag{5}$$

This is also called "simple random sampling." The assumption here is that if the sample is selected by a random sampling method, where all elements in the population have equal chance, data bias would not taint research. However, in real life, random sampling may not always be appropriate or available. The circumstances under which the data was collected, for instance a crime scene with limited sample availability, or in a situation where data accessibility is limited, random sampling may not be the best choice. In addition, one requirement of simple random sampling is that the population size must be known hence the term $\binom{N}{n}^{-1}$. However, if the population is dynamic, i.e. changing and the drop out of existing elements or addition of new elements changes and shifts population size, then simple random sample may not be possible. In most cases, *non-equal probability sampling* is the common option in sampling method of the Midzuno scheme.(Sampath, 2005; pp. 73-74; see also Midzuno, 1952; pp. 99-107) For instance, the non-equal probability sampling of the Midzuno scheme is given by:

$$P(sample) = \begin{cases} \dfrac{\hat{X}}{X} \dfrac{1}{\binom{N}{n}} \text{if } n(s) = n \\ 0 \qquad \text{otherwise} \end{cases} \tag{6}$$

Where $\hat{X}$ = unbiased estimator of population total: $\hat{X} = \frac{N}{n} \sum_{i \in s} X_i$; X = population total of the size variable x in random sampling (Sampath, pp. 73-74).

It may be argued that even if the population size is unknown (non-finite), it may be capable of being estimated. In Sampath, the estimation of the population, under the Horvitz-Thompson estimator method, is explained thus:

"Theorem 1.6 The Horvitz-Thompson estimator $\hat{Y}_{HT} = \sum_{i \in s} \frac{Y_i}{\pi_i}$ is unbiased for the population total and its variance is

$$\sum_{i=1}^{N} Y_i^2 \left[\frac{1-\pi_i}{\pi_i} \right] + 2 \sum_{i=1}^{N} \sum_{\substack{j=1 \\ i<j}}^{N} Y_i Y_j \left[\frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j} \right]$$

Proof The estimator $\hat{Y}_{HT} = \sum_{i=1}^{N} \frac{Y_i}{\pi_i} I_i(s)$

Taking the expectation of both sides we get

$$E_P[\hat{Y}_{HT}] = \sum_{i=1}^{N} \frac{Y_i}{\pi_i} \pi_i = Y$$

Therefore

$$\hat{Y}_{HT} - Y = \sum_{i=1}^{N} \frac{Y_i}{\pi_i} I_i(s) - \sum_{i=1}^{N} Y_i$$

$$= \sum_{i=1}^{N} \frac{Y_i}{\pi_i} [I_i(s) - \pi_i]$$

Squaring both the sides and taking expectation we get

$$E_P[\hat{Y}_{HT} - Y]^2 = \sum_{i=1}^{N}\left[\frac{Y_i}{\pi_i}\right]^2 E_P[I_i(s) - \pi_i]^2$$

$$+ 2\sum_{i=1}^{N}\sum_{\substack{j=1 \\ i<j}}^{N}\frac{Y_i}{\pi_j}\frac{Y_j}{\pi_j} E_p[I_i(s) - \pi_j]$$

$$= \sum_{i=1}^{N}\left[\frac{Y_i}{\pi_i}\right]^2 V_P[I_i(s)] + 2\sum_{i=1}^{N}\sum_{\substack{j=1 \\ i<j}}^{N}\frac{Y_i}{\pi_j}\frac{Y_j}{\pi_j}\mathrm{cov}_P(I_i(s), I_j(s))$$

$$= \sum_{i=1}^{N} Y_i^2\left[\frac{1-\pi_i}{\pi_i}\right] + 2\sum_{i=1}^{N}\sum_{\substack{j=1 \\ i<j}}^{N} Y_i Y_j\left[\frac{\pi_i - \pi_i\pi_j}{\pi_i\pi_j}\right]$$

Hence proof. □" (Samplath, pp. 14-19).

where the inclusion indicator was defined as (Sampath, p. 4):

$$I_i(s) = \begin{cases} I \ldots if \ldots s \in i, 1 \le i \le N \\ 0 otherwise \end{cases} \tag{7}$$

The inclusion probability is defined as:

$$\pi_i = \sum_{s \in i} P(s) \quad \text{and} \qquad \pi_{ij} \sum_{s \in i, j} P(s) \tag{8}$$

The "approach", i.e. Horvitz-Thompson method (Hovitz, 1952; pp. 663-6) as discussed in Sampath, is not helpful in obtaining the total population N . Although it provides series of steps in mathematical proof, it presents little practical value. Therefore, it is necessary to look for other practical methods to estimate the population size. Without the complicated approach displayed by Sampath, the Horvitz-Thompson method could be explained simply as a method to estimate the mean and total of a superpopulation in a stratified sample. This point (stratified sampling) was not properly explained by Sampath when the Horvits-thompson was introduced. The Horvits-Thompson was introduced as "the most popular … estimator for population total …" when in fact it should have been qualified as an estimator for population total in stratified sampling (*see* Sampath, pp. 4-8). In so doing, the inverse weighting is used to account for the different proportion found in each strata of the population. Let $Y_i = 1,2,...,n$ be an independent sample from n taken from the sample space of $N \ge n$ where each n_i is a distinct stratum with the common mean of μ . Under this method, it is further assumed that there is a probability that each random item in n_i will be included in each stratum is called inclusion probability defined as π_i. Under this set up, the Horvits-Thompson estimator is given by:

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \pi_i^{-1} Y_i \tag{9}$$

where $\hat{Y}$= estimated superpopulation; the subscript HT= Horvitz-Thompson; π= inclusion probability; and Y_i= independent sample, i.e. strata, $Y_i = 1,2,...,n$ taken from the sample space of $\Omega : N \geq n$.

From (3.6), the estimated mean is given by:

$$\hat{\mu}_{HT} = N^{-1}\hat{Y}_{HT} \quad (10)$$

Which means that $\hat{Y}_{HT}$ is a portion of the total population or $\hat{\mu}_{HT} = \frac{\hat{Y}_{HT}}{N}$. By substituting (3.6), statement (3.7) becomes:

$$\hat{\mu}_{HT} = \frac{Y_i}{N\pi_i} \quad (11)$$

The process by which the estimated mean μ_{HT} may be thought of as an estimate obtained through series of bootstrap or Jackknife resampling technique (*viz.* Roderick and Rubin, 2002).

There are several tools used in population estimates; these methods can be classified into two types: (i) direct and (ii) indirect methods for estimating population size. The direct method is exemplified by the Release-Recapture method (RR). The RR method is practical for non-human population, i.e. capture, release, and recapture animals in the while. For human population, this method may be impractical because it is not easy to "tag" humans. The RR method is defined as:

$$\hat{N} = \frac{\text{(Total captured)(Number marked)}}{\text{Total caputred with mark}} \quad (12)$$

This method is known as *direct method* because it involves the use of actual counting of the population. The estimate of the population size is based on these trials of capture, release, and recapture. In social science, this capture-release-recapture method may not be feasible because it is not easy to "mark" or "tag" human beings. However, in specific cases, such as in marketing research or tracking customers, the method may be a useful tool for determining customer base or market size. For instance, tracking customers in a department store by telephone numbers or membership card is an equivalent of what zoologist does in the wild by capture-release-recapture subjects.

The second direct method of estimating population is to work with known population size from the prior period and adjust it for the number of new births, deaths and net migration. This second method is given as:

$$P_{t+n} = P_t + B_{t,t+n} - D_{t,t+n} + M_{t,t+n} \quad (13)$$

where

P_{t+n} = change of population from one period to the next; P_t= population size of the last period; $B_{t,t+n}$ = live birth that occurs during time between t and $t+n$; $D_{t,t+n}$ = death during the interval; and $M_{t,t+n}$ = net migration, i.e. $= M_{t,t+n} = I_{t,t+n} + O_{t,t+n}$

For an *indirect method* of population estimation is the census method, the formula for the census method is given as:

$$P_{est} = P_1 + \frac{n}{N}\left(P_2 + P_1\right) \quad (14)$$

where P_{est} = population estimate; n = number of months from P_1 census date of the estimate; N = number of months between census period; P_2 = population size in the last census; and P_1 = population in the second to last census.

From the two methods shown above, population size is not easily determined. Under the first method, capture-tag-release-recapture may not be feasible in human population study unless the population is in a controlled environment, such as club card scheme used in supermarket to track customers. The second method relies on prior population estimate. This method is feasible only if prior population size is known. In random sampling, it is necessary that the population size be known because random sampling is an equal probability sampling method. This sapling method requires that the population size be known in order to determine the probability of each element in the population. When the population size is not known or is uncertain, then equal probability sampling may not be feasible. Therefore, random sampling is equally not feasible in most circumstances. Moreover, random sampling serves one purposes: preventing selection biased. As for randomness testing, it is the data not the sampling method that matters.

The objective of the tests for randomness in this writing are not concerned with sampling methodology, but are concerned on the data itself. The query is "whether the data comes from a random process?" these two issues: random sample and random sampling in relations to data classification is summarized in the table below.

The difficulty presented in this second method is the use of the prior population size. When the population is dynamic, this number can only be an estimate. If time is allowed to lapse and the counting be affected, it is possible to count the population or determine the population size. However, without the benefit of a prior count, the population size may be difficult to determine. In some cases, researchers are faced with unknown population size (non-finite population).

Table 3: Data Produced by Random and non-Random Sampling Methods

	DATA SAMPLE		**Nota Bene**
SAMPLING	Random	Non-Random	*Random sampling does not guarantee a random data.*
Random	*RR*	*RN*	
Non-Random	*NR*	*NN*	

The ideal result of using random sampling method to obtained random data (*RR*) may not be achievable. Even if the sampling method is random, if the data itself comes from a population that is not random, then RN will be produced. For example, if the survey asks about a sensitive issue where members of the public hold different opinions on the issue, i.e. "Do you believe in capital punishment? In this case, the dispersion of the opinion may be wildly and is unpredictable; thus, the data is random (*RR*). Does it mean that the research is good? No, it means that the question solicits divisive responses.

Under the same sampling method: simple random sampling, the data may show non-randomness. For example, in public opinion survey, the selection method may be a simple random selection. However, if the population holds the same or similar opinion on a particular issue, i.e. "Should there is a free and fair election?" The answer would most likely be {Yes}. In this case, the data is not random. Random by a common definition is lacking predictability. In this case, there is predictability. Does it mean that the research is bad? No, it just means that the data is not a random number set.

Other two scenarios: *NR* and *NN* may also occur. These different types of data results are not indicative of the quality of the research. Therefore, the use of sampling method as the lead indicator for testing randomness is not advisable. The focal point of the examination should be the data itself. Answer the query" "did the data come from a random process?" is more interesting than asking: "was simple random sampling method used?" The focus of this test is on the testing for "randomness in data" not in the sampling method.

4.0 UNIVARIATE DATA

4.1 Randomness in univariate case

In univariate case, the data would have to be codified into binary data form, i.e. 1 and 0 in a (1, 0) string. Univariate case draws on the battery of random tests offered by NIST. NIST battery of tests uses (1, 0) data string is the starting point to verify randomization. In cases where the data is in a form of categorical data, i.e. (Yes | No), the conversion into (1, 0) is a straight forward matter, i.e. (Yes = 1, No = 0). This type of binary data follows binomial distribution. The point of analysis may begin with the Laplace rule of Success to determine pointwise probability. The success is equated to Yes and Yes is equated to 1; the category No = 0. The probability of success is given by:

$$P(s) = \frac{s+1}{N+2} \tag{15}$$

where s = number of success or the counts of score of 1, and N = total number of observations. The probability of non-success or failure is given by:

$$P(f) = 1 - p \tag{16}$$

By convention, the probability of success is denoted as p and the probability of failure (non-success) is denoted as q. The pointwise probability forecast is given by:

$$P(X) = \frac{n!}{(n-X)!X!} p^n q^{n-X} \tag{17}$$

The test statistics follows the Z-equation, thus:

$$Z_{bin} = \frac{\frac{X}{n} - p}{\sqrt{\frac{pq}{n}}} \tag{18}$$

Assume that the data set is comprised of an answer choice in a scalar form: (0, 1, 2, 3). Generally, this form of answer score card is treated as a continuous data and, thus, continuous probability applies. However, in this case, the scale will be treated as dichotomous data (discrete) by dividing the score into zero and non-zero. The zero (0) will be counted as zero and the non-zero scores will be counted as 1. Thus, the scale is converted thus: $(0,1,2,3) \to (0,[1,1,1,])$. For example, if there are five questions where the answer were: (2, 3, 2, 0,0). The score of non-zeroes were (2,3,2). These non-zeroes are converted to (1,1,1). Now the new data is in the form of (1,0) where there are through counts of 1 and two counts of 0 or (1,1,1,0,0). With this data form, the value *p* may be determined, thus:

$$P(s) = \frac{s+1}{N+2} = \frac{3+1}{5+2} = \frac{4}{7} = 0.57 \quad \text{and} \quad P(f) = 1 - p = 1 - 0.57 = 43 \text{ or}$$

$$p = 0.57 \text{ and } q = 0.43 .$$

With known *p* and *q*, the test statistics for any give number of *X* may be determined. Assume that one wants to predict the probability of $X = 3$, the probability be determined, thus:

$$P(3)=\frac{n!}{(n-X)!X!}p^{n}q^{n-X}=\frac{3!}{(5-3)!3!}0.57^{5}0.43^{5-3}=\frac{6}{2(6)}0.57^{5}0.43^{2}$$

$$=\frac{6}{12}0.57^{5}0.43^{2}=0.50(0.06)(0.18)$$

$$=0.50(0.011)$$

$$=0.0055$$

The test for statistical significance follows:

$$Z_{bin}=\frac{\frac{X}{n}-p}{\sqrt{\frac{pq}{n}}}=\frac{\frac{3}{5}-0.57}{\sqrt{\frac{0.57(0.43)}{5}}}=\frac{0.60-0.57}{\sqrt{\frac{0.2451}{5}}}=\frac{0.03}{\sqrt{0.4902}}=\frac{0.03}{0.2214}=0.1355$$

where $Z(0.1355)$ is equal to 0.556 or 55.60% chance of happening. The question then remains: Is this data string (1,1,1,0,0) random number or does the data set (1,1,1,0,0) come from a random process? Or does the number come from a random selection?

The question deals with two completely different things. "A number or data coming from a random process" focuses on the data itself. That the data comes from a random process is an inference made after having completed a random test, i.e. NIST or other forms of tests to verify whether the data are random numbers. The null hypothesis is generally stated that the number comes from a random process, if there is a statistical significance finding, it is said that the number did not come from a random process. The rationale for this posture of the null hypothesis is that the term "randomness" has a functional equivalence of equal probability of elements distributed within the sampling space. Thus, if all elements of the population in the sampling space have equal probability of being selected, then it is normally distributed within that space. Within such a space, if the confidence interval is set at, say 0.95 or 0.99, any elements of the data found outside of this predetermined C.I., it may then be concluded that they are non-random because if they are random, they would have been fallen within the C.I. region under the distribution curve.

The second part of the question asks whether the data "comes from a random selection." This second part of the question may not have a definite answer because a selection method or sampling method does not determine the type of data being selected. For instance, if the process under which the sample has been draw was random, i.e. the data was randomly distributed a sampling method will not change this nature of the data. That is to say if the data is randomly distributed, a non-random sampling method would still produce a random data and a random sampling method would also produce random data. The sampling method is not a determining factor. It is the nature of the data *res ipsa* that defines whether the data is random.

4.2 Monost bit test

Monobit test is the first of the 15 binary test for randomness offered by NISt. It is based on frequency count. In this test, there is one sequence or one block where $M=1$ consisting of data string of 1 and 0. The objective is to compare the string of 1 to the string of 0 and determine whether there is a sensible pattern. If there is a pattern, it means that the sequence is not random. Since the sequence is comprised of 1 and 0, it can be model by a one-dimension random walk. Assume that p is the probability of 1 and $q=1-p$. The total probability is given by $p+q=1$. Asume further that the taking of steps walking to the right is n_1 and the steps taken to the left is n_2. Thus, the total steps taken would be $n_1+n_2=N$. Therefore, the probability of taking n_1 steps to the right is given by:

The monobit test is a test based on binomial distribution because it deals with binary data: (1, 0) which falls within the category of Bernoulli random variables. The notation used for even occurrence is X_i; the string of these event which is called data string may be written as: $X_i = \varepsilon$ where $X = 2\varepsilon - 1$. The sequence is given as $S_n = X_1, X_2, ..., X_n = 2(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n) - n$. In binary data analysis, it is necessary to designate the event of interest or category of interest. In the monobit test, the category is interest is 1. The probability of 1 in any time sequence is ½. If the number of observations is sufficient large, according to De Moivre-Laplace theorem, (Feller, 1968; vol. 1, sect. VII.3) the distribution of the binomial sum will approximate normal distribution. Recall that the Moivre-Laplace theorem is given by:

$$\lim_{n\to\infty} P_n\left[a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right] = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^1/2}dx \qquad (19)$$

where $a, b \in R \cup \{\pm\infty\}$ and $a < b$ because a is the beginning point and b is he ending point in the integration. The convergence in a and b is uniform.

If $y \in \square, n \in \square$ and the probability of the event lies between 0 and 1: $0 < p < 1$; this condition defines:

$$k(y) = \left\lfloor np + y\sqrt{np(1-p)} \right\rfloor \qquad (20)$$

Therefore,

$$\lim_{x\to\infty} \sum_{j=0}^{k(y)} \binom{n}{j} p^j (1-p)^{n-j} = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-c}dx \qquad (21)$$

where …

$a \quad = -\infty$

$b \quad = np + y\sqrt{np(1-p)}$

$c = \dfrac{(z-np)^2}{2np(1-p)}$

The monobit test is based on the Central Limit Theorem for a random walk: $S_n = X_1, X_2, ..., X_n$. The CLT is given by:

$$\lim_{n\to\infty} P\left(\frac{S_n}{\sqrt{n}} \le z\right) = \Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\Delta}^{z} e^{-u^2/2}du \qquad (22)$$

If (z) is positive, the probability of the random walk is given by:

$$P\left(\frac{|S_n|}{\sqrt{n}} \le z\right) = 2\Phi(z) - 1 \qquad (23)$$

4.3 Frequency (Monobit) test

The frequency test attempts to verify whether the data string is random or came from a random process by looking at the ratio of 1 and 0 to the entire data string.

$$S_n = \sum_{i=1}^{n} X_i \tag{24}$$

where $X_i = (a+b)$ and that $a = \sum 1_i$, $b = \sum -1_i$ and $-1 = 0$. In the example, the string consists of (1,1,1,0,0); thus, $a = \sum 1_i = 1+1+1 = 3$ and $-1 = 0 \rightarrow -1, -1$ because there are two counts of zero. Therefore, $b = \sum -1_i = -2$. Now, the value for S_n may be easily determined thus: $S_n = 3 + (-2) = 1$. With known S_n , the observed ratio of 1 may now be determined: S_{obs} which given by:

$$|S_{obs}| = \frac{S_n}{\sqrt{n}} \tag{25}$$

The value for S_{obs} is $S_{obs} = 1/\sqrt{5} = 1/2.24 = 0.4464$.

The test statistic of the frequency or monobit test is given by:

$$Z = \frac{S_{obs}}{\sqrt{2}} \tag{26}$$

From known S_{obs} the calculation for $Z_{monobit}$ follows: $Z = \frac{S_{obs}}{\sqrt{2}} = \frac{0.4464}{1.4142} = 0.3157$. Looking at the Z-table, find the corresponding p-value for $Z(0.3157)$, the value is 0.626. Under 0.95 confidence interval, this number is within the 0.95 confidence interval; there is no statistical significance. Recall that the hypothesis formulation under the frequency or monobit test is:

$$H_0 : (1-\alpha) < 1.65\} \rightarrow Random$$
$$H_A : (1-\alpha) > 1.65\} \rightarrow Non-Random$$

The null hypothesis is based on: $s = |S_n|/n$. This is compared to the observed value: $|s(obs)| = |= X_1, X_2, +... + X_n|.\sqrt{n}$. The p-value or the area under the curve, bounded by the lower and upper bounds of the confidence interval (CI), is given by:

$$pValue = 2[1 - \Phi(|s(obs)|)] = erfc\left(|s(ob)|/\sqrt{n}\right) \tag{27}$$

The notation *erfc* stands for the *complementary error function* which is given by:

$$erfc(z) = \frac{2}{\sqrt{\pi}} \int_{z}^{\infty} e^{-u^2} du \tag{28}$$

In this case, $1-\alpha = 0.626$ which is less than 1.65. Therefore, H_0 (null hypothesis or the assumption of randomness) cannot be rejected or the number string (1,1,1,0,0) came from a random process. Note that the position of the null hypothesis is assuming that the data comes from a random process. The alternative hypothesis is to prove otherwise. This is *counter intuitive* if one wants to prove randomness and the assumption of the alternative hypothesis is "non-randomness." NIST recommends that the sample size for this test is $n = 100$. In the example above, only five observations were used.

4.4 Frequency test within block

In the first case of frequency test, the entire string of data is treated as one block (M); thus, $n = M$. In the second case of frequency test: frequency test within block, the data string is segmented into *M* blocks; each *M* has peculiar patterns which is illustrated by the *Aperiodic Templates* below for a small value of pattern *m* where $2 \leq m \leq 5$.

$m = 2$	$m = 3$	$m = 4$	$m = 5$
01	001	0001	00001
10	011	0011	00011
	100	0111	00101
	110	1000	01011
		1100	00111
		1110	01111
			11100
			11010
			10100
			11000
			11110

Source: Barbour, and Holst, L., and Janson, S. (1992). Poisson Approximation. Oxford: Clarendon Press. Sects. 8.4 and 10.4. *Cited in* Rukhin A., Soto J., Nechvata J., Smid M., Barker E., Leigh S., Levenson M., Vangel M., Banks D., Heckert A., Drat J. and Vo S. (2010). *A Statistical Test Suite for Random and Pseudo-Randomg Number Generators for Cryptographic Applications*. National Institute of Standards and Technology (NIST); technology Administration, U.S. Department of Commerce. Special Publication 800-22, rev. 1a by Lawrence E. Bassham III, (April 2010). p.3-10.

The "frequency test within block" is still dealing with binary digit (bit). A bit is defined as a block of data that conveys information. Generally, the block is comprised of values of 1 and 0. A block of (1, 0) combination is known as a bit. Eight bits equals one byte. Hence, kilobyte (1,000 x 8 = 8,000 bits), Megabytes (1,000,000 8 = 8,000,000 bits), and gigabytes (100,000,000 x 8 = 8,000,000,000 bits) are referring to the information in binary form of a unit length of eight bits.

The first frequency test under NIST scheme is called monobit test because it involves only one block of bit. Each block is called M-bit block. In the frequency test, the block is a single block or the block size is $M = 1$. In "frequency test within block," the M-bit block is more than 1, i.e. $M > 1$. The purpose of the frequency test within block is to verify whether the frequency of ones within the M-bit block is approximately $M/2$ because the assumption of randomness assumes that the frequency of ones is approximately $M/2$. The following definitions are provided:

ε = bit sequence generated by the random number generator (RNG) or pseudo-random number generator (PRNG);
M = bit block; and
n = length of the bit string.

The test statistic used for verification is the chi-square test. The chi-square measures the ratio of the frequency of ones observed and comparing it to the expected ratio---which is approximately $M/2$ or half of the bit block. The chi-square test statistic is given by:

$$\chi^2 = \sum_{i=0}^{K} \frac{(v_i - N\pi_i)^2}{N\pi_i} \tag{29}$$

Alternatively, equation (2.1) is simplified to:

$$\chi^2 = 4M \sum_{i=1}^{N} \left(\pi - \frac{1}{2} \right)^2 \tag{30}$$

where the term π_i is defined as:

$$\pi_i = \frac{\sum_{j=1}^{M} \varepsilon_{(i-1)M+j}}{M} \tag{31}$$

The p-value for the test is determined by:

$$Pvalue = \int_{\chi^2_{obs}}^{\infty} \frac{e^{-u/2} u^{N/2-1}}{\Gamma(N/2) 2^{N/2}} du = \int_{\chi^2_{obs/2}}^{\infty} \frac{e^{-u} u^{N/2-1} du9}{\Gamma(N/2)} = igamc\left(\frac{K}{2}, \frac{\chi^2_{obs}}{2} \right) \tag{32}$$

This is known as *incomplete gamma function*.

Assume that the following string is given: $\varepsilon = 0110011010$. The length of the string is 10; therefore, $n = 10$. The *first step* is to divide the string into non-overlapping block is called N. The non-overlapping block is defined as $N = \left\lfloor \frac{n}{M} \right\rfloor$. Therefore, the string $\varepsilon = 0110011010$ can be divided into blocks thus:

011 = first N
001 = second N
101 = third N
0 = discarded.

The objective is to divide the data string into blocks. Each block (M) must contain at least one 1 in relations to relative strong of zeroes. How many ones should be in a block? Notice that in the string $\varepsilon = 0110011010$, the separating point for the block is the ending of 1 that would make the shortest block pattern in the string; as for the first data of the next block, it may be 1 o 0. Therefore, $N = 3$. Note that each block is 3 or $M = 3$; thus, the number of block is $N = \left\lfloor \frac{n}{M} \right\rfloor = \left\lfloor \frac{10}{3} \right\rfloor = \lfloor 3.33 \rfloor = 3$. The 0.33 is disregarded as excess incomplete block. This first step is called *partition input sequence*.

Second step is to determine the proportion of ones in each M-bit block. This proportion is denotes as π_i which is given by equation (5.3):

$\pi_i = \frac{\sum_{j=1}^{M} \varepsilon_{(i-1)M+j}}{M}$ for $1 \le i \le N$ which reads $\pi_i = \frac{\sum_{j=1}^{M} \varepsilon_{(i-1)M+j}}{M} = \frac{Ones}{\text{bit block}}$, from the blocks:

011	=	first N	has	2 ones
001	=	second N	has	1 one
101	=	third N	has	2 ones
0	=	discarded.		

Therefore, π_i includes:

$\pi_1 = 2/3$

$\pi_2 = 1/3$

$\pi_3 = 2/3$

Third step, using equation (5.2.), the value of χ^2_{obs} my be determined, thus:

$$\chi^2 = 4M\sum_{i=1}^{N}\left(\pi - \frac{1}{2}\right)^2$$
$$= 4(3)\left[\left(\frac{2}{3}-\frac{1}{2}\right)^2 + \left(\frac{1}{3}-\frac{1}{2}\right)^2 + \left(\frac{2}{3}-\frac{1}{2}\right)^2\right]$$
$$= 12[(0.66-0.50)^2 + (0.33-0.50)^2 + (0.66-0.50)^2]$$
$$= 12[(-0.16)^2 + (-0.17)^2 + (-0.16)^2]$$
$$= 12[(0.0256) + (0.0289) + (0.0256)]$$
$$= 12(0.0801) = 0.9612$$

Fourth step, find the critical value for chi-square in the chi-square table at a given degree of freedom and percentage confidence level. The degrees of freedom in the chi-square equation is $df = N = 3$.

For 0.99 confidence interval, the chi-square critical value is 11.30 and for 0.95 confidence interval, the critical value is 7.80. The value for the observation is $\chi^2_{obs} = 0.9621$ which is less than both 7.80 for CI of 95% and less than 11.30 for CI of 99%. Recall that the hypothesis formulation for the random test is:

$H_0 : \chi^2_{obs} < \chi^2_{0.95} = random$

$H_A : \chi^2_{obs} > \chi^2_{0.95} = non - random$

In this case, it is not possible to reject the null hypothesis. Therefore, it is concluded that the data string $\varepsilon = 0110011010$ comes from a random process.

Equation (30) and equation (31) use K and M degrees of freedom respectively. The following Tables provides the corresponding classes of data string with possible combination of probabilities.

Table 4: *K = 3; M = 8*

Classes	**Probabilities**
$v \leq 1$	$\pi_0 = 0.2148$
$v = 2$	$\pi_1 = 0.3672$
$v = 3$	$\pi_2 = 0.2305$
$v \geq 4$	$\pi_3 = 0.1875$

Table 5: *K = 5; M = 128*

Classes	**Probabilities**
$v \leq 4$	$\pi_0 = 0.1174$
$v = 5$	$\pi_1 = 0.2430$
$v = 6$	$\pi_2 = 0.2493$
$v = 7$	$\pi_3 = 0.1752$
$v = 8$	$\pi_4 = 0.1027$
$v \geq 9$	$\pi_5 = 0.1124$

Table 6: *K = 5; M = 512*

Classes	**Probabilities**
$v \leq 6$	$\pi_0 = 0.1170$
$v = 7$	$\pi_1 = 0.0.2460$
$v = 8$	$\pi_2 = 0.2523$
$v = 9$	$\pi_3 = 0.1755$
$v = 10$	$\pi_4 = 0.1027$
$v \geq 11$	$\pi_5 = 0.1124$

Table 7: *K = 5; M = 100*

Classes	**Probabilities**
$v \leq 7$	$\pi_0 = 0.1307$
$v = 8$	$\pi_1 = 0.2437$
$v = 9$	$\pi_2 = 0.2452$
$v = 10$	$\pi_3 = 0.1714$
$v = 11$	$\pi_4 = 0.1002$
$v \geq 12$	$\pi_5 = 0.1088$

Table 8: *K = 6; M = 10,000*

Classes	**Probabilities**

$v \le 10$	$\pi_0 = 0.0882$
$v = 11$	$\pi_1 = 0.2092$
$v = 12$	$\pi_2 = 0.2483$
Les 1 - $v = 13$	$\pi_3 = 0.1933$
$v = 14$	$\pi_4 = 0.1208$
$v = 15$	$\pi_5 = 0.0675$
$v \ge 16$	$\pi_6 = 0.0727$

Source: Tables 1 – 5 *see* Pal Revesz (1990). Random Walk in Random and Non-Random Environments, Singapore: World Scientifc. Pp. 55; *cited in* Rukhin et al. p. 3-5.

4.5 Run test for randomness

The Runs Test is considered non-parametric because it does not depend on any parameter as in a polynomial equation. The test examines the substrings of consecutive 1's and 0's. These strings are considered homogeneous. The objective is to look at the oscillation of the substring and verify whether the oscillation is too fast or too slow. The following terms are defined:

V_n = number of runs;

π = fixed proportion, i.e. $\pi = \sum_j \varepsilon_j / n$

The reference point is 0.50; the objective of the test is to determine whether the substring is close to 0.50 or $\left|\pi - \frac{1}{2}\right| \le \frac{2}{\sqrt{n}}$ which is defined by the probability of the runs as:

$$\lim_{x\to\infty} P\left(\frac{V_n - 2n\pi(1-\pi)}{2\sqrt{n}\left(\pi(1-\pi)\right)}\right) \tag{33}$$

In order to evaluate the run (V_n), define that $k = 1, 2, ..., n-1$ and $r(k) = 0$ if the present string is equal to the next string, i.e. $\varepsilon_k = \varepsilon_{k+1}$ and $r(k) = 1$ if $\varepsilon_k \ne \varepsilon_{k+1}$. Then the number of runs is given by:

$$V_n = \sum_{k=1}^{n-1} r(k-1) \tag{34}$$

The pValue or the area under the curve within which is considered non-significance region or the region bounded by the lower and upper boundaries of confidence interval is given by:

$$pValue = erfc\left(\frac{\left|V_n(obs) - 2n\pi(1-\pi)\right|}{2\sqrt{2}\left(\pi(1-\pi)\right)}\right) \tag{35}$$

The value of the runs V_n explains the characteristic of the oscillation. If V_n is too large, it means that the oscillation is fast and if the value of V_n is small, it means that the oscillation us slow. The significance tests centers around the observation and test of this oscillation.

The run test (also known as Wald-Wolferwitz test) (Wald and Wolfowitz, 1940; pp. 147-162; Mendenhall *et al.*, 1986) can be used to verify if the data comes from a random process

(Bradley, 1068; chap. 12). A "run' is series of increasing values or series of decreasing values; a run is a ranked data set. A ranked data set is a data set arranged in ascending values or descending values. The number of data values in the set is called the "length" of the run. In random data test, the probability that $(l+1)$ is larger or smaller than l^{th} follows binomial distribution---the basis of the runs test. For example, the following is a data set that has been categorized into positive (+) and negative (-) signs:

+ + + + + - - - - - - - + + + - - - - + + + + - - - - - - -
1 2 3 4 5 6

There are 6 runs or $R = 6$ and the length of the data string is 30 counts. The run is counted by counting the times that the sign change from positive (+) to negative (-). In this case, there are 6 alternations. The purpose of the test is to verify that each data point is mutually independent of the others in the set. This mutual independence is an evidence of randomness. If the data in the set are not mutually independent, it means that one event depends on the other or that one event may be used to predict the occurrence of the other, such a condition is said *not* mutually independent and, therefore, is not random. The argument of the null hypothesis is that the data set is random; the alternative hypothesis is that the data is not random.

H_0 : the sequence was produced in random manner
H_1 : the sequence was NOT produced in a random manner

The test statistic for the runs test is given by:

$$Z = \frac{R - \bar{R}}{S_R} \tag{36}$$

where R = number of observed number of runs; $\bar{R}$ = expected number of runs:

$$\bar{R} = \frac{2n_1 n_2}{n_1 + n_2} + 1 \tag{37}$$

S_R = standard deviation of the runs where …

$$S_R^2 = \frac{2n_1 n_2 (2n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)} \tag{38}$$

In the alternative, the variance may be written as:

$$\sigma^2 = \frac{(\mu - 1)(\mu - 2)}{N - 1} \tag{39}$$

Recall that $\mu = \bar{R}$ and $S_R^2 = \sigma^2$ and n_1 = number of + and n_2 = number of – event.

The significance level is *alpha*: α . The runs test rejects the null hypothesis if:

$$|Z| > Z_{1-\alpha/2} \tag{40}$$

For large sample is defined as $n_1 > 10$ and $n_2 > 10$. For a large sample, use $Z_{0.95} = 1.65$. The runs test can answer the following question: *Were the sample data generated from a random process?*

For a numerical illustration, the following data set is provided in a pre-classified (=) and (-) form:

\+ + + + + - - - - - - - + + + - - - - + + + + - - - - - - -
1 2 3 4 5 6

The objective is to use the Runs Test to verify whether the data string comes from a random process. In this string, there are 6 runs and the size of the data set or the length of the data string is 30 counts or $n = 30$. The counts are divided into two categories: (-) and (-). The next step is to determine the correct value for $n_1 + n_2 = 30$. Assume that $n_1 = +$ counts and $n_1 = -$ counts. Segments 1, 3 and 5 carry (+) signs. There are a total of 12 counts (+) and 18 counts of (-). Therefore, $n_1 = 12$ and $n_2 = 18$.

The second step is to determine the value for $\bar{R}$, thus:

$$\bar{R} = \frac{2n_1n_2}{n_1+n_2} + 1 = \frac{2(12)(18)}{12+18} + 1 = \frac{216}{30} + 1 = 7.2 + 1$$
$$\bar{R} = 8.20$$

The third step is to calculate the standard deviation, by using the variance formula of the Runs test, thus:

$$S_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)} = \frac{2(12(18)((2(12)(18)) - 12 - 18)}{(12+18)^2(12+18+1)} = \frac{432(432-12-18)}{900(31)}$$
$$= \frac{432(402)}{27900}$$
$$= \frac{173,664}{27900}$$
$$= 6.2245$$

Recall that standard deviation is the square root of the variance; therefore, the standard deviation for this Runs test is: $S_R = \sqrt{S_R^2} = 2.50$. The test statistic may now be calculated, thus:

$$Z = \frac{|R - \bar{R}|}{S_R} = \frac{6 - 8.20}{2.50} = \frac{2.20}{2.50} = 0.88$$

With the critical value of 0.88, the percentage probability may be looked up in the Z-table. For $Z(0.88)$, the p-value is $1 - \alpha = 0.811$ which is less than 0.95. Recall that the hypothesis statements were:

$$H_0 : Z_{obs} < Z(0.95) \rightarrow Random$$
$$H_A : Z_{obs} \geq Z(0.95) \rightarrow Non-Random$$

The critical value for $Z(0.95)$ is 1.65. In this case, the critical value for $Z(0.88)$ is 0.811. Since $Z(0.88) < Z(0.95)$ or $0.811 < 1.65$. The null hypothesis cannot be rejected. Therefore, it is concluded that the data came from a random process.

Run-sequence plots are easy way to graphically summarize a univariate data set. Common assumptions of univariate data set is that they behave like: (i) random drawing; (ii) come from a fixed distribution; (iii) with common location; and (iv)with common scale. The run-sequence allows us to see the shift in location and scale. It also allows us to detect the outliers. The run-sequence is formed by the vertical axis (response variable Y_i) and the horizontal axis ($X_i(1,2,...)$). The questions answered by the run-sequence are: (1) Are there any shifts in location? (2) Are there any shifts in variation? And (3) Are there any outliers?

IV. CONCLUSION

The concept of randomness is the foundation of statistics since statistical test deals with distribution and percentage probability of event occurrence within the distribution. In both types of probability: discrete and continuous, each event in the distribution space is considered random event, hence, equal probability distribution in normal distribution cases. Therefore, since randomness plays an indispensable role in the discussion of statistical tests, it is necessary for researcher to know how to test for randomness in the data. This paper presents three direct tests for randomness in discrete data. In addition, we also present three indirect tests fort randomness through trend analysis. The three direct tests were under NIST approach. The proposed three indirect tests were a mixed of (i) reverse Arrangement Test (RAT), (ii) Military Handbook Test (MHT), and (iii) La Place Trend test (LTT).

REFERENCES

Bradley (1968). Distribution-Free Statistical tests, Chapter. 12.

Feller, W. (1968). An Introduction to Probability Theory and Its Applications (Volume 1). Wiley. ISBN 0-471-25708-7. Section VII. 3.

Horvitz, D. G.; Thompson, D. J. (1952). “A generalization of sampling without replacement from a finite universe.” Journal of the American Statistical Association, 47, 663–685.

Perry, John (1953). The Story of Standards, Funk and Wagnalls, Library of Congress Cat. No. 55-11094, p. 123.

Mendenhall, Scheaffer, and Wackerly (1986), Mathematical Statistics with Applications, 3rd Ed., Duxbury Press, CA.

Midzuno, H. (1952).“On the sampling system with probability proportionate to sum of sizes,” Ann. Inst. Statist. Math., 3 (1952), pp. 99–107.

Roderick J.A. Little, Donald B. Rubin (2002). Statistical Analysis With Missing Data, 2nd ed., Wiley. ISBN 0-471-18386-5.

Sampath, S. (2005). Sampling Theory and Methods. 2nd ed. Alpha Science International. Harrow: UK. ISBN 1-84265-214-1., pp. 73-74.

Wald, A. and Wolfowitz, J. (1940). “On a test whether two samples are from the same population. Ann. Math Statist. 11, 147-162.

OFFICE OF THE EDITOR-IN-CHIEF
International Journal of Research & Methodology in Social Science (IJRMSS)
420 Bourichane Rd., Unit 1 Ban Naxay
Xaysetha, Vientiane LAOS
Tel. +856 21 263585 ☆
Editor@socialsciencepublication.com
EditorIJRMSS@gmail.com
www.socialsciencepublication.com

JOURNAL

The International Journal of Research and Methodology in Social Science (IJRMSS) is a double blind peer-review, open access, academic research, published quarterly. The journal publishes research articles, literature review articles, technical commentaries, and case studies.

MISSION

The mission of the journal is to provide publication service to the academic and research community in social science that is: (i) peer reviewed, (ii) high exposure, (iii) true open-access, (iv) fully achieved and retrievable online, and (v) leads to impact factor score.

OBJECTIVE

The objective of the journal is to provide an open access online forum and resource for researchers to share, disseminate, and exchange ideas and research outcome.

POLICY

It is the policy of the journal to provide high quality research publication that is a true open access journal. The journal is committed to "no fees – no charge" to authors. All articles will be subjected to blind peer review process. We are committed to high quality research publication with free online access and strictly observe non-predatory publication policy.

CALL FOR PAPERS

IJRMSS is published quarterly. Our call for papers is on-going. Authors may submit the research article(s) by email to: Editor@socialsciencepublication.com or EditorIJRMSS@gmail.com. We welcome articles in the following fields and their interdisciplinary

SOCIAL SCIENCE
Anthropology
Archaeology
Criminology
Demography
Economics
Geography (human)
History
International relations
Law
Linguistics
Pedagogy
Political science
Psychology
Science education
Sociology

INTERDISCIPLINARY
Applied physics
Artificial intelligence
Bioethics
Bioinformatics
Biomedical engineering
Biostatistics
Cognitive science
Complex systems
Computational linguistics
Cultural studies
Cybernetics
Environmental science
Environmental social science
Environmental studies
Ethnic studies
Evolutionary psychology
Forestry
Library science
Mathematical & Theoretical biology
Mathematical physics
Medicine
Military science
Network science
Neural engineering
Neuroscience
Science studies
Scientific modeling
Semiotics
Sociobiology
Statistics
Systems science
Urban planning
Web science

All papers submitted must conform to the following format:

Font: Times New Roman 12 points
Margin: Single line spacing; 1 inch top, bottom, right & left.
Reference: APA style. https://owl.english.purdue.edu/owl/resource/560/01/

File: MS Words. Originality will be checked by TURNITIN.
Template: https://drive.google.com/file/d/0B2B2CwoNDenGejVCSlhMNzRMeUE/view?usp=sharing

COPYRIGHT

> "**17 U.S. Code § 107 - Limitations on exclusive rights: Fair Use**
> Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall include—
> (1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes;
> (2) the nature of the copyrighted work;
> (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
> (4) the effect of the use upon the potential market for or value of the copyrighted work.
> The fact that a work is unpublished shall not itself bar a finding of fair use if such finding is made upon consideration of all the above factors. *See* Pub. L. 94–553, title I, § 101, Oct. 19, 1976, 90 Stat. 2546; Pub. L. 101–650, title VI, § 607, Dec. 1, 1990, 104 Stat. 5132; Pub. L. 102–492, Oct. 24, 1992, 106 Stat. 3145."

ERRATA

In case where an error escapes the review process or is discovered after the publication has already been released, IJRMSS will contact the author for written comments and/or correction, and the journal will publish the correction in the *errata* section in the subsequent issue no later than 2 publication cycles after the error has been discover. Errata will be published when the error is material and may mislead or misrepresent facts. Minor error, such as typographical error or unless it involves mathematical formula, does not compel errata publication.

PLAGIARISM

IJRMSS maintains strict policy on originality of works and proper credit given to prior sources. We employ computer software to check for work originality. The article found more than 10% similarity with prior works will be automatically rejected for cause and no invitation for resubmission would be issued. Originality check does not count instances of citation and repeated citation of names, titles, and/or sentences and paragraphs in quotation or in-text quotation when proper quotation marks are used.

DISCLAIMER

The opinions expressed in each article are those of the individual author(s). Specific questions about the article should be directed to the corresponding author(s). The contact information of the corresponding author is listed in the footnote on the first page of the article marked with a star: ★. IJRMSS has editorial discretion to decline articles that are demeaning or disrespectful or otherwise intended to cause public outcry in matters of race, national origin, religion, ethnicity, and/or sexual orientation where public sensitivity outweighs the academic value of the proposed article(s).

INTERNATIONAL JOURNAL OF RESEARCH & METHODOLOGY IN SOCIAL SCIENCE

VOL. 4, NO. 1	JAN. – MAR., 2018	QUARTERLY JOURNAL	ISSN 2415-0371

OPEN ACCESS

Berlin Declaration of Open Access to Knowledge in the Science and Humanities
Budapest Open Access Initiatives
The Cape Town Open Education Declaration

INTERNATIONAL JOURNAL OF RESEARCH & METHODOLOGY IN SOCIAL SCIENCE
ISSN 2415 – 0371

OFFICE OF THE EDITOR-IN-CHIEF
Email: Editor@socialsciencepublication.com

420 Bourichane Rd., Unit 1 Ban Naxay
Xayasetha, Vientiane, LAOS
Tel. +856 21 263585
Email: EditorIJRMSS@gmail.com

55 / 123 Rangsit-Nakhon Nayok Rd.
T. Lamlookka, A. Thanyaburi
Pathumtanee, THAILAND
Email: EditorIJRMSS@gmail.com

www.socialsciencepublication.com