2016 UKSim-AMSS 18th International Conference on Computer Modelling and Simulation

Accessibility of Norwegian Municipalities Websites:

An Interactive Learning Environment Experimental Investigation

Ahmed A. Abdelgawad, Jaziar Radianti and Mikael H. Snaprud
Department of ICT,
University of Agder
Grimstad, Norway
{ahmedg|jaziar.radianti|mikael.snaprud}@uia.no

John Krogstie
Department of Computer and Information Science,
Norwegian University of Science and Technology
Trondheim, Norway
krogstie@idi.ntnu.no

***Abstract*—Accessibility is an important aspect of websites generally and public websites particularly. Many ways could be proposed to enhance accessibility, however the impact of selected actions is hard to predict due to diversification and contradiction, in addition to the existence of the time factor. A System Dynamics simulation model including factors affecting the accessibility of Norwegian municipal websites was encapsulated in an Interactive Learning Environment (ILE). As the model promised to be able to change how users think and take decisions, this ILE was tested by users in an experiment. We have conducted α, β, and γ change analysis on the results of this experiment. Results showed that the ILE was successful in changing 50% of its users' understanding and perceptions about the system's causal relationships and policy options, and helping 30% redefining the standards they use to assess or evaluate these relationships and policy options.**

Keywords-Experimental Design; Alpha, Beta, Gamma Analysis; Accessibility; System Dynamics; Municipal Websites; Interactive Learning Environment

I. INTRODUCTION

Accessibility of a website refers to the ability of all people to use a website irrespective of their disabilities or the client devices they use to access internet [1]. Accessibility is an important aspect of websites in general and of public websites in particular, to be able to serve all citizens equally. For example if a webpage is designed to retrieve user input solely via mouse clicks, people with disabilities preventing them from using a mouse or people using mobile phones to access internet will not be able to use this webpage.

Accessibility of a website can be assessed in terms of compliance with a set of accessibility metrics defined by guidelines like WCAG 2.0 [1] and ISO 9241-20 [2]. It may be evaluated quantitatively using methodologies like UWEM [3]. Many researches and projects have addressed evaluation of public websites accessibility, for instance [4], [5]. In addition, governments carry out benchmarking and pass laws to increase public websites accessibility. In Norway, the Agency for Public Management and eGovernment (DIFI) used to evaluate governmental websites annually [6]. For a long period, accessibility was one of three indicators used to be measured by DIFI to encourage governmental agencies to increase the accessibility of their websites.

A public website, as an Information System, consists of people (like web-masters, editors, developers, etc.) and procedures, besides telecommunications, hardware, software, and data [7]. When considering the management process and the people component of eGovernment website, many ways could be proposed to enhance its accessibility like consulting experts, replace CMS, recruiting, and training. The expected impact of these measures ranges from slow to fast, cheap to expensive, and short-term to long-term. This diversification and contradiction, in addition to the existence of the time factor make the decision of adopting only one way to do the task a challenge. More challenging is how to prioritise limited resources to achieve the best effect on accessibility. A simulation model is a very efficient tool that can be employed to deal with such situation [8].

Based on the results retrieved form a set of in-depth interviews with web-masters and editors from different Norwegian municipalities, Abdelgawad, Snaprud, and Krogstie [9] identified various factors and causal relationships governing the processes having an impact on accessibility of Norwegian municipal websites, and compiled these relationships into a System Dynamics (SD) simulation model titled "Accessibility of Norwegian Municipalities Websites".[1] This model is intended to work as a decision support tool by helping eGovernment websites' managers to take informed decisions, and decision makers to find policies that enable governmental organisations to enhance their websites' accessibility.

From another angle, the model is supposed to be able to change how its users think and take decisions. It is allegedly capable of changing their understanding and perceptions about the system's causal relationships and policy options, in other words changing their mental models [8]. In this paper, the major problem and consequently research question we are interested in answering is whether this model is really

[1] The model is licensed under a Creative Commons Attribution Share Alike license, and available at: http://forio.com/simulate/ahmedg/accessibility-of-norwegian-municipalities-websites-a-decision-support-tool/model/

978-1-5090-0888-9/16 $31.00 © 2016 IEEE
DOI 10.1109/UKSim.2016.16

capable of changing its users' understanding and perceptions about the system's causal relationships and policy options as it promises or not. If the model is capable of doing this, we can expect that by using it, the websites' managers and decision makers will be able to take more informed decisions. To answer our research question, we have developed an Interactive Learning Environment (ILE) [8] to be an interface for the model, prepared a survey tool, and conducted an experiment with users to understand the effect of using the ILE and accordingly the model on them.

The remainder of this paper is organised as follows: the next section will describe the architecture of the ILE developed for this research. In addition, it will provide a detailed description of the procedure followed to conduct the experiment, including the analysis method. The section that follows will explore and discuss the results of the experiment. The last section concludes the paper.

II. Research Methodology

A. *ILE Architecture*

1) *Client Side*

The ILE we have developed for this experiment is web-based. It was built using Forio's Epicenter. [2] Its Graphical User Interface (GUI) has 4 tabs: Home, Instructions, Control Panel, and Dashboard. The Home tab gives a brief introduction to the topic of the ILE, including basic knowledge about website's accessibility and policy options. The Instructions tab puts the user in the context of using the ILE, including specific instructions to guide her/him through the simulation or game-play. The Control Panel tab has all policy options provided by the ILE to control the simulation, in addition to simulation time progress buttons.

The simulation starts at year 0 and can be progressed year by year or to the end of the simulation at year 6. In the Control Panel, the user can reset the simulation and start a new scenario from the beginning, whether the current scenario reached year 6 or not. Policy options available are represented by graphical control elements for managing workforce, managing workforce time, training workforce, consulting vendor and upgrading website technology (CMS). The Dashboard tab has charts showing over time behaviour of important simulation variables, needed by the user to stand on the results reflected by her/his policies entered to the Control Panel.[3]

Epicenter is a very powerful tool, having all what is needed to build an ILE, nevertheless for our interface charts, we have replaced Forio's Polymer-based[4] charts with our own JavaScript charts. Our JavaScript code for charts is still based on Forio's code, and uses the same powerful open source Forio's Contour,[5] but in addition, it is capable of showing many scenarios on the same chart.[6] This way the users are able to compare different scenarios.

2) *Server Side*

The ILE is fully functioning by using solely the client-side, yet we wanted to log users' interactions with the ILE i.e. record users' decisions and results. To achieve this, Epicenter uses Node.js[7] for client server communications, which then could be logged to a database; however this is limited to paid subscribers. We wanted to have a generic architecture that could be used by everyone.

To log users' interactions, we have developed JavaScript snippets[8] and added them to all decision control elements and charts available on the GUI. These JavaScript snippets communicate with a PHP file called "forioepicenter.php".[9] We developed this file to save the values sent by the GUI to MySQL database.[10] Furthermore, a survey tool was needed to deploy our pre- and post-test questionnaires. We opted for an open source PHP-based tool called Limesurvey.[11] Fig. 1 shows the ILE architecture.

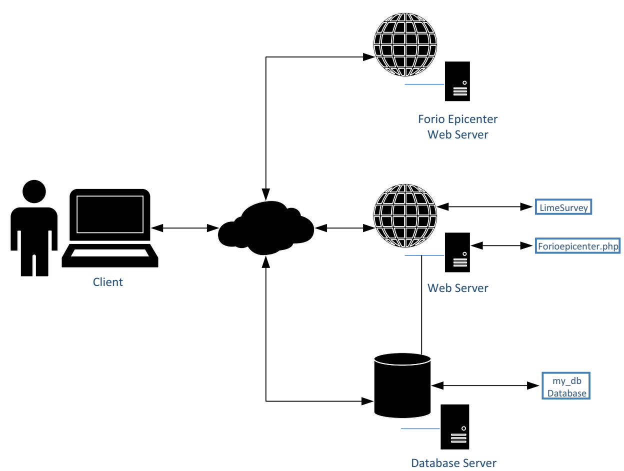

Figure 1: ILE Architecture

[2] Forio's 3rd generation platform for simulation, modelling, and analytics. Available at: http://forio.com/

[3] The ILE is available at: https://forio.com/app/ahmedg/eaccessibility/eaccessibility.html

[4] https://www.polymer-project.org

[5] https://github.com/forio/contour

[6] Our JavaScript code for charts is generic, so that others can use it in building their ILEs. The JavaScript file is available at: https://forio.com/app/ahmedg/eaccessibility/elements/contour-chart.js

[7] https://nodejs.org

[8] Our code snippets are available inside the HTML of the ILE. It could be shown by viewing the page source. We have marked these snippets by HTML comment tags "<!--begin " and "<!--end ", and made them generic to be easily copied to any other ILE.

[9] forioepicenter.php can be deployed in any server/web-hotel supporting PHP. We made forioepicenter.php available at: https://forio.com/app/ahmedg/eaccessibility/helper/forioepicenter.php

[10] MySQL database tables needed by forioepicenter.php, can be reproduced at any MySQL using my_db.sql which we made available at: https://forio.com/app/ahmedg/eaccessibility/helper/my_db.sql

[11] https://www.limesurvey.org

B. Procedure

We have conducted our experiment with volunteer students at the University of Agder, Norway, on the 8th of September 2015. A couple of weeks earlier we started spreading the invitation for a game-play session with free pizza in Grimstad campus of the university. In the day of the experiment 17 students showed up. Some faced technical troubles with the experimentation system, and by the end we could extract 12 useful finished surveys. Properties of the participants are presented in Table I.

TABLE I: PARTICIPANTS WHOSE RESULTS WERE ACCEPTED

Property	Value	%
Age Group	18-24	67%
	25-34	25%
	35-44	8%
Gender	Male	67%
	Female	33%
Field of Study/Work (specialisation)	ICT	92%
	Mechatronics	8%
Knowledge of Math Modelling	Yes	25%
	No	75%
Knowledge of System Thinking/Dynamics	Yes	17%
	No	83%

The experiment session took 1 hour. For 20 minutes, the experiment supervisor gave a presentation to introduce the topic and the ILE to the participants. The presentation included the terms the participants would experience during the intervention using the ILE. By the end of the presentation, the participants were asked to connect to the Limesurvey server prepared earlier via their web-browsers. In addition to the free pizza that was promised to everyone, the 2 highest-performing participants were promised a piece of Egyptian pharaonic collectable each.

The testing session started for everyone by answering the pre-test questionnaire. The pre-test questionnaire consists of 10 Likert 5-point scale items (strongly disagree, disagree, neutral, agree, or strongly agree). These items were statements about the dynamics of the system in terms of causal relationships and possible policy options. All these statements are designed to test the participants' knowledge about the system and possible policy options. For example, participants were asked to report their level of agreement or disagreement with this statement "*Upgrading CMS takes long time to show an effect on the value of website accessibility*".

The Pre-test questionnaire was supposed to take no more than 5 minutes; nevertheless it was left to the participants to take as much time as they need. The participants were informed that they can ask the supervisor for help at any time; however we abstained from providing any help that could lead to biases in their answers.

The intervention using the ILE or the game-play started as the participant ended the Pre-test questionnaire, without the option of going back to the Pre-test questionnaire. The game-play was limited to 25 minutes. Afterwards all participants were directed automatically to the post-test questionnaire, without the option of going back to the intervention session.

The Post-test questionnaire contained exactly the same Likert items used in the Pre-test questionnaire, however after answering how she/he thinks now about each statement after the game-play, the participant was asked to think back in time before the game-play, and report how much she/he agreed or disagreed with the same statement based on her/his new understanding. This is called the Retrospective Pre-test or the Then-test [10]. It is very common that participants change their understanding between the Pre-test and the Post-test [11]. The Then-test gives the participant the opportunity to re-answer the Pre-test based on her/his new understanding/ perception after the intervention. In this case, the Post-test and the Then-test have the same base frame of reference [11].

1) α, β, and γ Change

In 1976 Golembiewski *et al.* [12] distinguished among 3 different types of change as a result of an intervention, namely α, β, and γ. α change refers to a real change, i.e. a change in the reading on a fixed measurement scale. β change marks a measurement scale redefinition in terms of changing the intervals of the measurement scale itself, i.e. a measurement scale recalibration. γ change reflects to a conceptual change or re-conceptualisation, i.e. a redefinition of the measurement construct [12], [13]. Many methods to assess α, β, and γ changes appeared since 1976, including the method suggested by Golembiewski and his colleagues. According to a comprehensive literature review conducted by Riordan *et al.* [13], there are 5 major methods to detect α, β, and γ changes:

- Ahmavaara's technique [12]
- Actual-ideal difference measures [14]
- Retrospective accounts [15]
- Confirmatory factor analysis [16]
- Latent growth modelling [17]

Our expectation of the number of the participants volunteering to our experiment was very modest, because of limited participants' availability as well as financial support. Based on that, we have opted for using the Retrospective accounts method, as it is the only method that doesn't require a large sample, in addition to that it can test for α, β, and γ change independently [13]. It is worth mentioning that although we are not aware of any application of this method in assessing SD ILE effect, using this specific method for that purpose was suggested by Friedman, Cavaleri, and Raphael [18].

2) Retrospective Accounts

As the Post-test and the Then-test questionnaires are answered based on the same understanding/perception as mentioned above; the Retrospective accounts method detects α change by detecting the change between the Post-test and the Then-test. Furthermore, since the Pre-test and the Then-test are basically measuring the same thing based on either 2 different understandings/perceptions or 2 differently calibrated measurement scales, γ and β change are detected by detecting the change between the Pre-test and the Then-test [15].

Even though the Retrospective accounts method supports analysis on both group and individual levels, we have chosen to focus merely on the individual level analysis, because of the limited number of participants. After all, group change is the sum of its individuals' change. Sometimes certain individual change could be overlooked by detecting only group changes [19], [20]. Furthermore, "*a large amount of change exhibited by only a few individuals may be taken as evidence that the intervention had a group effect*" [20].

To apply the Retrospective accounts method to our data, we have followed the practice of Birkenbach [19] in general. Nevertheless, we have opted for following Brodersen and Thornton [20] in detecting γ change first, then remove the participants showing γ change from the process of detecting α and β. According to Porras and Singh [21], when γ change is detected, α and β detection becomes problematic.

Answers to the questionnaire items from the Pre-, the Post-, and the Then-tests of each participant were used as raw data/basic data points [21]. Consequently for every participant, we have compiled 3 paired samples Pre, Post, and Then. The first step in the analysis is to try detecting γ change per participant. Terborg and his colleagues [15] suggested 2 methods:

1. Using Correlation

For every participant, correlations between Pre & Then ($r_{\text{Pre Then}}$), Post & Pre ($r_{\text{Post Pre}}$), and Post & Then ($r_{\text{Post Then}}$) are calculated. To test for differences between $r_{\text{Post Then}}$ & $r_{\text{Pre Then}}$ and $r_{\text{Post Then}}$ & $r_{\text{Post Pre}}$, Williams's test to compare correlations of 2 paired/dependant samples is used to calculate $t_{(r_{\text{Post Then}})(r_{\text{Pre Then}})}$ and $t_{(r_{\text{Post Then}})(r_{\text{Pre Post}})}$ respectively [20]. γ change exists if the following conditions are met:

a) $r_{\text{Post Then}}$ is substantially greater than $r_{\text{Post Pre}}$

b) $r_{\text{Post Then}}$ is substantially greater than $r_{\text{Pre Then}}$

2. Using Standard Deviation

Standard Deviations for Pre (s_{Pre}), Post (s_{Post}), and Then (s_{Then}) are calculated for every participant. Morgan-Pitman test to compare variances of 2 paired/dependant samples is used to calculate $t_{(s_{\text{Post}})(s_{\text{Then}})}$, $t_{(s_{\text{Pre}})(s_{\text{Then}})}$, and $t_{(s_{\text{Pre}})(s_{\text{Post}})}$. γ change exists if the following conditions are met:

a) s_{Post} is not different from s_{Then}

b) s_{Post} is different from s_{Pre}

c) s_{Then} is different from s_{Pre}

The highest level of γ change happens when both correlation and standard deviation methods to detect γ change occur concurrently [15]. If a participant doesn't show any signs of γ change, we start detecting β or α change.

To test for β or α change, mean values of Pre, Post, and Then are calculated for every participant, yielding $\overline{x}_{\text{Pre}}$, $\overline{x}_{\text{Post}}$, and $\overline{x}_{\text{Then}}$ respectively. Student's t-Test to compare means of 2 paired/dependant samples is used to calculate $t_{(\text{Then})(\text{Pre})}$, $t_{(\text{Then})(\text{Post})}$, and $t_{(\text{Post})(\text{Pre})}$. If $t_{(\text{Then})(\text{Post})} > t_{(\text{Then})(\text{Pre})}$, descriptively speaking, then there is more evidence of α change than β change, and vice versa [15].

Terborg and his colleagues [15] emphasised on that t-statistics on the individual level analysis should in general be judged descriptively. Although the tests used to compute these statistics are for dependant/paired samples, which is the case, the inter-independency or independency condition inside each participant's Pre, Post, and Then samples is not met. Simply, inside each of them all data points come from the same participant [15].

III. Results and Discussion

Participant P6 and P11 were removed from the analysis because of showing no variance in their Pre, Post, or Then samples. We started the analysis by detecting γ change. This is done via correlation and standard deviation comparisons as mentioned above. The left half of Table II shows the needed correlation values in addition to the t-statistics calculated to compare them. The t-statistic columns at the left half of the table prove that $r_{\text{Post Then}}$ is substantially greater than both $r_{\text{Post Pre}}$ and $r_{\text{Pre Then}}$ for participants P2, P4, P5, P9, and P12, consequently detecting γ change. The t-statistics columns on the right half of the same table cannot at all prove that s_{Post} is not different from s_{Then}, while both are different from s_{Pre} for any participant, and consequently no γ change was detected based on standard deviation.

After eliminating participants showing γ change, from β and α change detection procedure, Table III shows that P1, P3, and P7 have smaller values of $t_{(\text{Then})(\text{Post})}$ compared to $t_{(\text{Then})(\text{Pre})}$ denoting β change for these participants. Accordingly participants P8, and P10 have exhibited α change. Table IV shows the overall α, β, and γ changes detected for all participants in comparison to their answers about mathematical modelling and system dynamics knowledge, in addition to their demographic data. 50% of the participants who were included the analysis have shown γ change, reflecting a change in their understanding and perceptions about the system's causal relationships and policy options. 30% have redefined/recalibrated the standards they use to assess or evaluate these causal relationships and policy options exhibiting β change. In total, 80% of the participants have redefined certain knowledge as a result of using the ILE, achieving the ILE's intended goals.

From an internal validity [22] point of view, to minimise testing validity threat, we have kept the questionnaires as merely Likert-scale items, and emphasised that there is no right or wrong answers, and participants needed to report what they think/believe. Furthermore, we made sure that all participants have fully understood questionnaire items since the pre-test, to account for any miss-understanding that could be automatically clarified during the post-test solely because of repetition. The same questionnaire was administered during pre- and post-test sessions to account for any instrumentation validity threat. Moreover, to eliminate experimenter bias, we have chosen self-report questionnaire type, and kept the whole experiment computerised without any human rater interactions, except when help to clarify any vagueness was needed.

To account for possible history validity threat, participants were asked to report their prior knowledge of mathematical modelling and system thinking/dynamics. Furthermore, the experiment time was limited to almost 1 hour, eliminating maturation or mortality validity threats. We have to admit that the research suffered from selection validity threat due to the availability of participants as previously mentioned. Nevertheless, this was somehow mitigated by the fact that participation was totally voluntary.

From external validity [23] perspective, participants were few, and limited to university students, yet they are mostly ICT students, who are expected –to some extent– to fill positions like website managers and decision-makers in the future, which are the users' positions originally targeted by the model. Other experiments with different samples are necessary. Longer periods between the pre-test, the treatment, and the post-test should be examined. Other sets of questionnaire items describing the model's causal relationships and policy option should be used in other experiments.

IV. Conclusion

Accessibility is an important aspect of websites in general and public websites in particular. Many ways could be proposed to enhance accessibility; however the expected impact of selected actions is hard to predict due to diversification and contradiction, in addition to the existence of the time factor, which makes decision making a challenge. An SD model includes factors affecting accessibility of eGovernment websites was encapsulated in an ILE. The model is allegedly capable of changing its users' understanding and perceptions about the system's causal relationships and policy options, in other words changing their mental models. In this paper, the major problem and therefore the research question we were interested in answering was whether this model is really capable of changing its users' understanding and perceptions as it promises or not.

In an experimental setting, the ILE was tested with users. We have applied α, β, and γ change analysis on the individual level to the results of this experiment. The results were that the ILE/model was successful in changing its users' understanding and perceptions about the system's causal relationships and policy options 50% of the time, and helping them in redefining the standards they use to assess or evaluate these relationships and policy options 30% of the time. In total, 80% of ILE users have redefined certain knowledge as a result of using it, achieving the ILE's intended goals. Based on these results, we recommend using the ILE/model in educating websites managers and decision makers about their systems.

Applying the α, β, and γ change analysis following the method of Terborg and his colleagues to test the effect of using SD ILE was easy and straight forward. More experimentation with larger samples, ideally including control groups, over longer time spans, and longer questionnaires are a very promising and recommended next step.

References

[1] 'Introduction to Web Accessibility', W3C, Sep-2005. [Online]. Available: http://www.w3.org/WAI/intro/accessibility. [Accessed: 06-Oct-2011].

[2] 'ISO 9241-20:2008', ISO, 2008. [Online]. Available: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40727. [Accessed: 11-Oct-2011].

[3] A. Nietzio, C. Strobbe, and E. Velleman, 'The Unified Web Evaluation Methodology (UWEM) 1.2 for WCAG 1.0', in Computers Helping People with Special Needs, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds. Berlin Heidelberg: Springer-Verlag, 2008, pp. 394–401.

[4] M. H. Snaprud, M. G. Olsen, and F. Aslaksen, 'Automatic Benchmarking and Presentation of the First Results from the European Internet Accessibility Observatory', in Exploiting the Knowledge Economy: Issues, Applications and Case Studies, M. Cunningham, Ed. IOS Press, 2006.

[5] M. G. Olsen, A. Nietzio, M. Snaprud, and F. Fardal, 'Benchmarking and improving the quality of Norwegian municipality web sites', in Proceedings of the 5th Int'l Workshop on Automated Specification and Verification of Web Systems, Castle of Hagenberg, Austria, 2009.

[6] 'Norge.no-Kvalitet', DIFI, 2009. [Online]. Available: http://kvalitet.difi.no/resultat/. [Accessed: 04-Oct-2011].

[7] E. Oz, Management Information Systems. Cengage Learning, 2008.

[8] J. Sterman, Business Dynamics: Systems Thinking and Modeling for a Complex World. McGraw Hill Higher Education, 2000.

[9] A. A. Abdelgawad, M. Snaprud, and J. Krogstie, 'Accessibility of Norwegian Municipalities Websites: A Decision Support Tool', in UKSim 5th European Modelling Symposium on Computer Modelling and Simulation, D. Al-Dabass, A. Orsoni, A. Pantelous, G. Romero, and J. Felez, Eds. Madrid, Spain: IEEE Computer Society, 2011, pp. 225–230.

[10] G. S. Howard, 'Response-Shift Bias A Problem in Evaluating Interventions with Pre/Post Self-Reports', Eval Rev, vol. 4, no. 1, pp. 93–106, Feb. 1980.

[11] S. K. Rockwell and H. Kohn, 'Post-Then-Pre Evaluation', Journal of Extension, vol. 27, no. 2, 1989.

[12] R. T. Golembiewski, K. Billingsley, and S. Yeager, 'Measuring Change and Persistence in Human Affairs: Types of Change Generated by OD Designs', Journal of Applied Behavioral Science, vol. 12, no. 2, pp. 133–157, Apr. 1976.

[13] C. M. Riordan, H. A. Richardson, B. S. Schaffer, and R. J. Vandenberg, 'Alpha, beta, and gamma change: A review of past research with recommendations for new directions', in Equivalence in Measurement, L. L. Neider, Ed. IAP, 2001.

[14] R. W. Zmud and A. A. Armenakis, 'Understanding the Measurement of Change', ACAD MANAGE REV, vol. 3, no. 3, pp. 661–669, Jul. 1978.

[15] J. R. Terborg, G. S. Howard, and S. E. Maxwell, 'Evaluating Planned Organizational Change: A Method for Assessing Alpha, Beta, and Gamma Change', The Academy of Management Review, vol. 5, no. 1, pp. 109–121, Jan. 1980.

[16] N. Schmitt, 'The use of analysis of covariance structures to assess beta and gamma change', Multivariate Behavioral Research, vol. 17, no. 3, pp. 343–358, 1982.

[17] D. Chan, 'The Conceptualization and Analysis of Change Over Time: An Integrative Approach Incorporating Longitudinal Mean and Covariance Structures Analysis (LMACS) and Multiple Indicator

Latent Growth Modeling (MLGM)', Organizational Research Methods, vol. 1, no. 4, pp. 421–483, Oct. 1998.

[18] S. Friedman, S. Cavaleri, and M. Raphael, 'Evaluating Changes in Systems Thinking Capacity: A Methodology Based on Alpha, Beta, Gamma Analysis', in Proceedings of the 21st International Conference of the System Dynamics Society, New York City, New York, U.S.A., 2003.

[19] X. C. Birkenbach, 'Self-Report Evaluations of Training Effectiveness: Measuring Alpha, Beta, and Gamma Change', South African Journal of Psychology, vol. 16, no. 1, pp. 1–7, Mar. 1986.

[20] D. A. Brodersen and G. C. Thornton, 'An investigation of alpha, beta, and gamma change in developmental assessment center participants', Perf. Improvement Qrtly, vol. 24, no. 2, pp. 25–48, Jan. 2011.

[21] J. I. Porras and J. V. Singh, 'Alpha, beta, and gamma change in modelling-based organization development', J. Organiz. Behav., vol. 7, no. 1, pp. 9–24, Jan. 1986.

[22] D. T. Campbell and J. C. Stanley, Experimental and Quasi-Experimental Designs for Research, 1 edition. Belmont, CA: Wadsworth Publishing, 1963.

[23] G. H. Bracht and G. V. Glass, 'The External Validity of Experiments', American Educational Research Journal, vol. 5, no. 4, pp. 437–474, Nov. 1968.

TABLE II: γ CHANGE DETECTION

	$r_{\text{Post Then}}$	$r_{\text{Pre Post}}$	$r_{\text{Pre Then}}$	$t_{(r_{\text{Post Then}})(r_{\text{Pre Post}})}$	$t_{(r_{\text{Post Then}})(r_{\text{Pre Then}})}$	γ	s_{Post}	s_{Then}	s_{Pre}	$t_{(s_{\text{Post}})(s_{\text{Then}})}$	$t_{(s_{\text{Pre}})(s_{\text{Post}})}$	$t_{(s_{\text{Pre}})(s_{\text{Then}})}$	γ
P1	0.62^{*}	0.46	0.56^{*}	0.55	0.20	--	1.29	0.99	1.06	0.94	0.62	0.22	--
P2	-0.59^{*}	0.23	0.08	-2.10^{**}	-1.83^{**}	Yes	1.07	0.53	1.25	2.71^{**}	-0.44	2.77^{**}	--
P3	0.80^{***}	0.60^{*}	0.27	0.82	2.88^{***}	--	1.05	1.05	1.58	0.00	-1.47	1.22	--
P4	1^{***}	0.90^{***}	0.90^{***}	13.50^{***}	13.50^{***}	Yes	0.85	0.85	0.95	--	-0.70	0.70	--
P5	-0.36	0.15	0.43	-1.48^{*}	-2.05^{**}	Yes	0.84	0.88	1.20	-0.11	-1.02	1.00	--
P6^{X}	--	0.41	--	--	--	--	1.03	0.00	1.16	--	-0.36	--	--
P7	0.12	0.21	0.21	-0.19	-0.19	--	1.45	0.82	1.03	1.70	1.00	0.66	--
P8	0.73^{**}	0.85^{***}	0.65^{**}	-0.75	0.54	--	1.20	0.92	1.23	1.10	-0.14	1.10	--
P9	0.89^{***}	0.36	0.62^{**}	3.96^{***}	1.50^{*}	Yes	0.99	1.37	1.35	-2.01^{*}	-0.94	-0.05	--
P10	0.84^{***}	0.64^{**}	0.66^{**}	1.23	1.05	--	1.37	0.88	0.92	2.43^{**}	1.50	0.18	--
P11^{X}	--	--	-0.32	--	--	--	0.00	1.03	1.35	--	--	0.82	--
P12	0.70^{**}	0.18	0.08	1.37^{*}	1.75^{*}	Yes	1.62	0.92	0.92	2.37^{**}	1.72	0.00	--

* $P < 0.10$ ** $P < 0.05$ *** $P < 0.01$

^{X} Participant removed because of showing no variance in Pre, Post, or Then

TABLE III: α AND β CHANGE DETECTION

	$\overline{x}_{\text{Post}}$	$\overline{x}_{\text{Then}}$	$\overline{x}_{\text{Pre}}$	$t_{(\text{Post})(\text{Pre})}$	$t_{(\text{Then})(\text{Post})}$	$t_{(\text{Then})(\text{Pre})}$	β	α
P1	2.90	3.10	2.70	-0.51	-0.61	-1.31	Yes	--
P2^{G}	3.40	2.50	3.30	-0.22	1.96^{*}	1.92^{*}	--	--
P3	3.00	3.00	2.60	-1.00	0.00	-0.77	Yes	--
P4^{G}	3.50	3.50	3.30	-1.50	n/a	-1.50	--	--
P5^{G}	2.60	2.90	3.10	1.17	-0.67	0.56	--	--
P6^{X}	2.80	3.00	2.70	-0.26	-0.61	-0.82	--	--
P7	3.10	2.70	3.20	0.20	0.80	1.34	Yes	--
P8	2.90	3.20	3.20	1.41	-1.15	0.00	--	Yes
P9^{G}	3.10	3.10	2.60	-1.17	0.00	-1.34	--	--
P10	3.10	2.90	2.80	-0.90	0.80	-0.43	--	Yes
P11^{X}	3.00	2.80	2.50	-1.17	0.61	-0.49	--	--
P12^{G}	3.20	2.80	2.80	-0.74	1.08	0.00	--	--

* $P < 0.10$ ** $P < 0.05$ *** $P < 0.01$ – ^{G} γ change detected

^{X} Participant removed because of showing no variance in Pre, Post, or Then

TABLE IV: α, β, AND γ CHANGE RESULTS VS PARTICIPANTS' PROPERTIES

	Knowledge of Math Modelling	**Knowledge of System Thinking/Dynamics**	**Field of Study/Work (specialisation)**	**Age Group**	**Gender**	**γ**	**β**	**α**
P1	Yes	Yes	ICT	18-24	M	--	Yes	--
P2	--	--	ICT	18-24	M	Yes	--	--
P3	--	--	ICT	18-24	M	--	Yes	--
P4	--	--	ICT	35-44	F	Yes	--	--
P5	Yes	--	ICT	25-34	F	Yes	--	--
P6^{X}	--	--	ICT	25-34	M	--	--	--
P7	--	--	ICT	18-24	M	--	Yes	--
P8	--	--	ICT	18-24	M	--	--	Yes
P9	Yes	Yes	Mechatronics	18-24	M	Yes	--	--
P10	--	--	ICT	18-24	F	--	--	Yes
P11^{X}	--	--	ICT	18-24	F	--	--	--
P12	--	--	ICT	25-34	M	Yes	--	--

^{X} Participant removed because of showing no variance in Pre, Post, or Then