

# Disabled People from Welfare to Jobs: *An Interactive Learning Environment Experimental Investigation*

Ahmed A. Abdelgawad<sup>\*</sup>, Jaziar Radianti<sup>\*</sup>, Mikael H. Snaprud<sup>\*</sup>, and John Krogstie<sup>§</sup>

<sup>\*</sup> University of Agder  
Department of Information and Communication Technology  
Jon Lilletunsvet 9, NO-4898 Grimstad, Norway  
[ahmedg@uia.no](mailto:ahmedg@uia.no), [jaziar.radianti@uia.no](mailto:jaziar.radianti@uia.no), [mikael.snaprud@uia.no](mailto:mikael.snaprud@uia.no)

<sup>§</sup> Norwegian University of Science and Technology  
Department of Computer and Information Science  
Sem Sælandsvei 7-9, NO-7491 Trondheim, Norway  
[krogstie@idi.ntnu.no](mailto:krogstie@idi.ntnu.no)

**Abstract:** *A System Dynamics model focusing on enabling disabled people to move from welfare to work in Norway was encapsulated in an Interactive Learning Environment (ILE) to contribute in linking ordinary people with policymaking. The ILE and the model behind it are intended to enable people to get a better understanding of the policy options to inform better decisions, and ultimately change users' mental models. To explore how this ILE could change how users think and take decisions, the ILE was assessed by an expert opinion poll, and tested with users in 2 different experiments. To identify the type of change this ILE is capable of causing, we have conducted  $\alpha$ ,  $\beta$ , and  $\gamma$  change analysis on the results of these experiments. 67% of the sample of our expert opinion poll think that the ILE achieves the intended goals. 33% of the users who were included the analysis of both experiments have shown a change in their understanding and perceptions of the system's causalities and policy options. 38% have redefined the standards they use to assess or evaluate these causalities and policy options. In total, 71% of the users have redefined certain knowledge as a result of using the ILE.*

**Keywords:** Experimental Design; Expert Opinion Poll; Alpha, Beta, Gamma Analysis; Disabled People Employment; System Dynamics; Interactive Learning Environment

## Introduction

Despite repeated statements from successive governments in Norway, the employment rates for disabled people<sup>1</sup> remains unchanged since 2000, on average only around 44% of disabled people in working age are in employment (Statistics Norway 2016); the numbers even became stable at 43% for the last 3 years (Statistics Norway 2016). In Norway's neighbour countries, the case is different. In Denmark, around 56% of disabled people are in employment (Bengtsson 2011), and 53% in Sweden (Galis 2012).<sup>2</sup> These low rates of employment among disabled people comes even under the protection of anti-discrimination law and a high governmental spending on employment efforts compared to any other country (Tøssebro 2009).

The Norwegian government has introduced different vocational rehabilitation programmes to increase the employment probability of disabled people through re-education. However, these vocational rehabilitation programmes suffer from many problems, for example according to Westlie (2008), disabled people with the lowest work opportunity will benefit the most from the vocational rehabilitation, however they have the lowest opportunity to be included in such programmes. Another research by Røed and Raaum (2006) shows that although vocational rehabilitation programmes in Norway increase the employment probability of disabled people, the net effect is nearly zero because of the time elapsed during the training without looking for jobs.

On the contrary, the Norwegian telecommunication company *Telenor*<sup>3</sup> has introduced a very successful 2 years vocational rehabilitation programme called *Open Mind Programme* –previously known as *Handicapped-Programme*– (Telenor Group 2012; Skøien, Hem, and Tyrmi 2006). At least 75% of this programme participants have got permanent jobs after finishing the programme (Skøien, Hem, and Tyrmi 2006). Although the programme is on a very small scale,<sup>4</sup> the results are encouraging. Moreover, Skøien, Hem, and Tyrmi in (2006) via simple calculations verified a net positive effect on government budget resulted from employing trainees of the Open Mind programme till their retirement.

Inspired by the idea of generalising the Open Mind programme country-wide, Abdelgawad *et al.* (2012) built a System Dynamics (SD) model “Disabled People from Welfare to Jobs”,<sup>5</sup> intending to enable a better understanding of factors that can enable and encourage the disabled people to take up jobs, taking into consideration the associated costs and savings for individuals, employers, and nationally. Figure 5 shows the model’s subsystems diagram.

---

<sup>1</sup> We use the term “disabled people” as it is the term used by *Statistics Norway* (<http://www.ssb.no>). A more correct term is “people with disabilities”, as used by the *Convention on the Rights of Persons with Disabilities* (<http://www.un.org/disabilities/convention/conventionfull.shtml>).

<sup>2</sup> Such employment rates are generally based on different national-level surveys, with different designs, timings, surrounding economic conditions, and definitions of being employed (OECD 2003).

<sup>3</sup> <http://www.telenor.no>

<sup>4</sup> On average 8 to 10 persons pass through the Open Mind annually (Skøien, Hem, and Tyrmi 2006).

<sup>5</sup> The model is licensed under the creative commons, and available at: <http://forio.com/simulate/ahmedg/disabled-people-from-welfare-to-jobs-a-decision-support-tool/model/>

This model is intended to enable its users to get a better understanding of the system’s causalities and policy options to inform better decisions from one side. From the other side the model aims at changing its users’ perceptions and attitudes about the issue, in other words changing their mental model (Sterman 2000). The envisaged approach promises to contribute in linking ordinary people with policymaking and rendering the decision-making processes easier to comprehend, for an ultimate goal of supporting more targeted discussions among individual citizens, disability organisations, and policy makers.

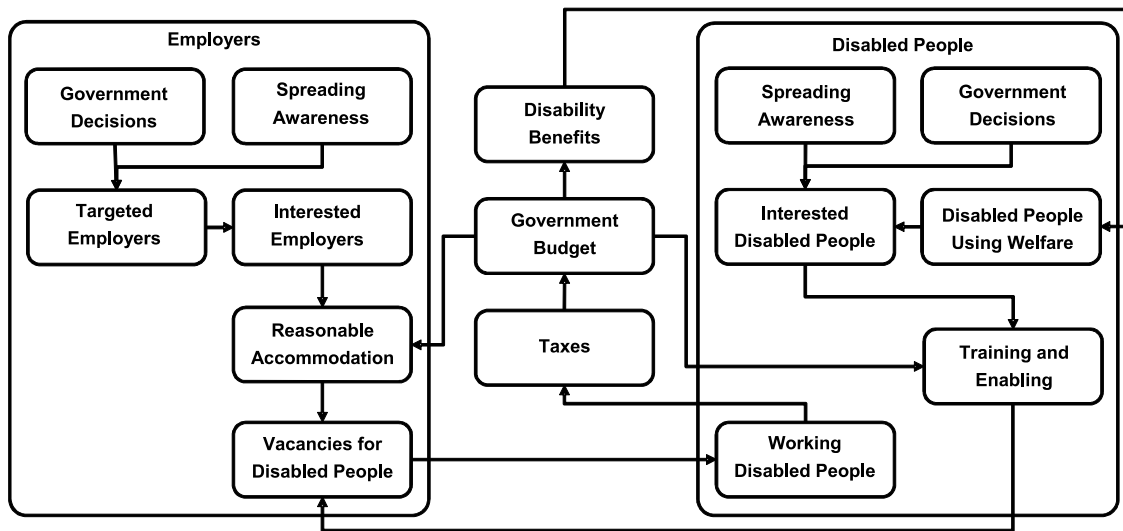


Figure 1. Subsystem diagram for the “Disabled People from Welfare to Jobs” model

The research question we pursue in this paper is to what extent this model is really capable of changing its users’ understanding and perceptions of the system’s underlying structure and policy options, and consequently achieve its goals.

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

## Research Methodology

To answer our research question, we have updated the model and developed an Interactive Learning Environment (ILE) (Sterman 2000) with the model in its core, prepared a testing tool, conducted an expert opinion poll, as well as 2 experiments with people using this ILE. The ILE consists of client-side used by users, and a server-side that was used to log users’ interactions with the system as well as answering the questionnaires of the experiments.

## ILE Framework

### Client-Side

The ILE we have developed for this experiment is web-based. It was built using Forio.com Epicenter.<sup>6</sup> We have used the best practices presented in Sterman (2014a; 2014b) to design our ILE's Graphical User Interface (GUI). The ILE GUI has 4 navigation tabs: Home, Instructions, Control Panel, and Dashboard, shown in Figure 2, Figure 3, Figure 4, and Figure 5 respectively. The Home tab briefly introduces the topic of the ILE, including basic knowledge about disabled people situation in Norway, Open Mind programme, and other policy options. The Instructions tab puts the user in the context of using the ILE, including specific instructions to guide her/him through the gameplay. The Control Panel tab has all policy options available by the ILE to control the simulation, in addition to simulation time progress buttons, which is either 5 years ahead or to the end of the simulation time.

The simulation starts in the year 2001 and can be progressed up to the year 2050. Nonetheless, the user can control the simulation merely starting from the year 2015. The period from 2001 to 2015 is included in the simulation solely to show the user a comparison between the behaviour of the model and the historical data. In the Control Panel, the user can reset the simulation and start a new scenario from the beginning, whether the current scenario reached the year 2050 or not. Policy options available are represented by graphical control elements to operationalise Open Mind-like programme country-wide including financing awareness campaigns targeting selected ratios of disabled people and potential employers, selected disabled people's preferences, and selected employers' preferences. The Dashboard tab includes charts showing over time behaviour of important simulation variables, needed by the user to stand on the current results reflected by her/his policies entered in the Control Panel.<sup>7</sup>

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<sup>8</sup> charts with our JavaScript charts. Our JavaScript code for charts is still based on Forio's charts code, and uses the same powerful open source Forio's Contour Library;<sup>9</sup> but in addition it is able to show many scenarios on the same chart. Our JavaScript code for charts is generic, so that others can use it in building their ILEs.<sup>10</sup>

---

<sup>6</sup> <http://forio.com>

<sup>7</sup> The ILE is available at: <https://forio.com/app/ahmedg/eparticipation/eparticipation.html>

<sup>8</sup> <https://www.polymer-project.org>

<sup>9</sup> <https://github.com/forio/contour>

<sup>10</sup> <https://forio.com/app/ahmedg/eparticipation/elements/contour-chart.js>

 Home

 Instructions

 Control Panel

 Dashboard


---

In Norway, the employment rates for disabled people is only 44% and remains unchanged since 2000. The government has introduced different vocational rehabilitation programmes to increase the employment probability of disabled people through re education. However, these programmes suffer from many problems.

Telenor has introduced a very successful 2 years vocational rehabilitation programme called Open Mind Programme. Open Mind consists of 3 months of class training, and 21 months of on the job training in Telenor or one of its partner companies. This approach combines on the job training, with a wage subsidised programme, that were both proven to be the most effective in increasing employability of disabled people. At least 75% of the trainees have got permanent jobs after finishing the programme. Although the programme is on a very small scale, the results are encouraging. Simple calculations were made to show the overall effect on government budget resulted from employing trainees of Open Mind till their retirement. It has been shown that accumulated net savings in terms of saved disability benefits and increased tax revenue after removing Open Mind programme cost is positive.

Inspired by the success of the Open Mind, we are interested in exploring factors that can enable more disabled people to take up jobs. To do that, we built a decision support tool to simulate the impact of possible government decisions over the next 35 years. The goal is enabling a better understanding of factors that can enable and encourage disabled people to take up jobs. In addition, the model takes into consideration the associated costs and savings for individuals, employers, and nationally through NAV and tax authority.

---

Next step: Select the  Instructions tab from the upper tab list, and read carefully.

**Figure 2: Home tab**

Imagine that you are a decision maker who is responsible for creating a better situation for disabled people, and ideally save public spending. The government can take decisions on spending related to disabled people and employers. You need to consider that your choices have to meet budget constraints and several preferences from disabled people and employers as well.

How to get started?

1. Select the 📄 Control Panel tab from the upper tab list to enter your decisions.
2. Make your decisions for the budget under the Government policies section.
3. Set the preferences both for disabled people and employers selecting either "yes" and "no" for each item.
4. After entering your decisions and preferences, you can run the simulation for only five years ahead by pressing "Progress 5 years" button, or to the end of simulation by pressing "Progress to 2050" button. You will be automatically transferred to the 📊 Dashboard to see the results of these decisions.
5. At any time during the simulation, you can press "Reset simulation" button in the 📄 Control Panel to restart the simulation, and start a new scenario again from 2015.
6. To see the results of your decisions at any time, select the 📊 Dashboard tab from the upper tab list.
7. In the 📊 Dashboard , you can press "Legend/Scenario Selector" button to select scenarios that will appear on charts.


Good luck in helping your people and managing the budget wisely!


**Figure 3: Instructions tab**

Current year: 2015

### Government Policies

Open Mind like Programme Start  
 Yes  
 No

Percentage of Disabled People Targeted by Awareness Campaign  
 100%

Percentage of Employers Targeted by Awareness Campaign  
 100%

### Disabled People Preferences

Combine Salary and Welfare Benefits  
 Yes  
 No

PC with Assistive Technology or Assistant  
 Yes  
 No

Reduce Administrative Barriers to Return to Welfare Support  
 Yes  
 No

Accessible Transportation to Work  
 Yes  
 No

### Employers Preferences

Recruit Disabled People for Public Procurement  
 Yes  
 No

Government Financial Support  
 Yes  
 No

Assistive Technology from Government  
 Yes  
 No

Government Financial Support to Adapt Workplace  
 Yes  
 No

Reduce Expected Overhead in Dealing with Social Benefits  
 Yes  
 No

Progress 5 years

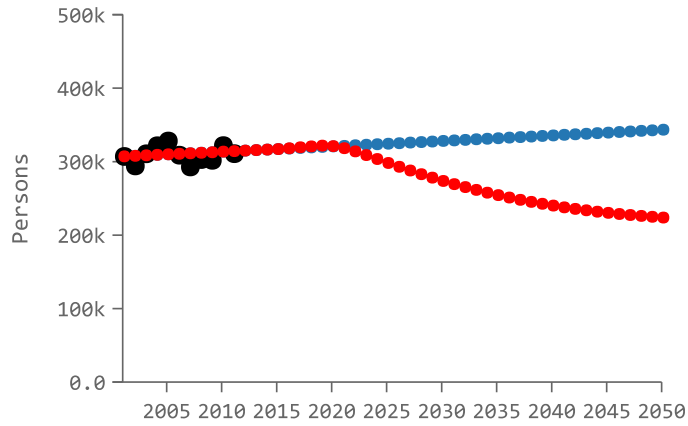
Progress to 2050

Reset simulation

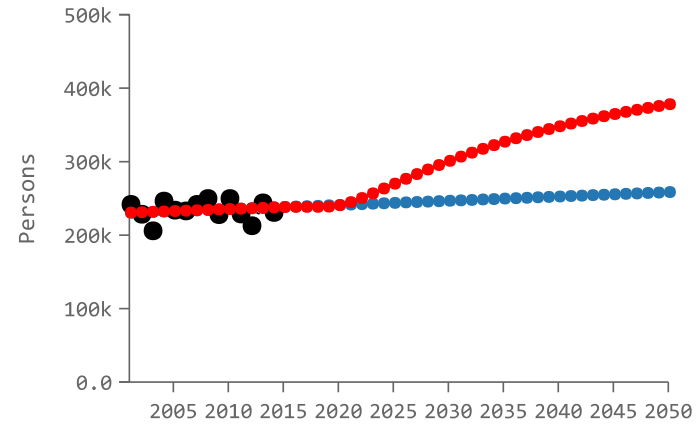
Figure 4: Control panel tab

Current year: 2050

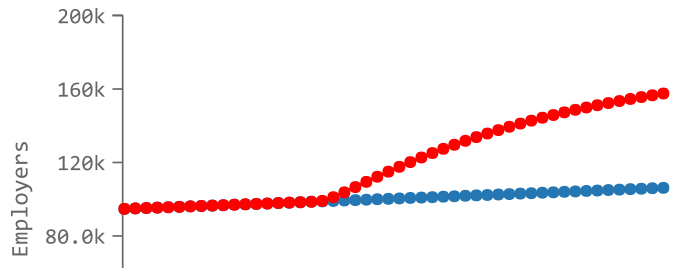
All Disabled People in Working Age, Using Welfare



Employed Disabled People



Employers Recruiting Disabled People



All Disabled People in Working Age

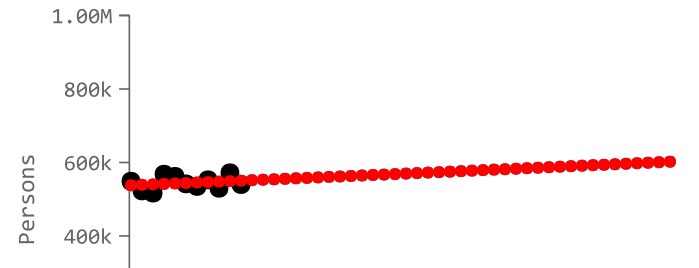
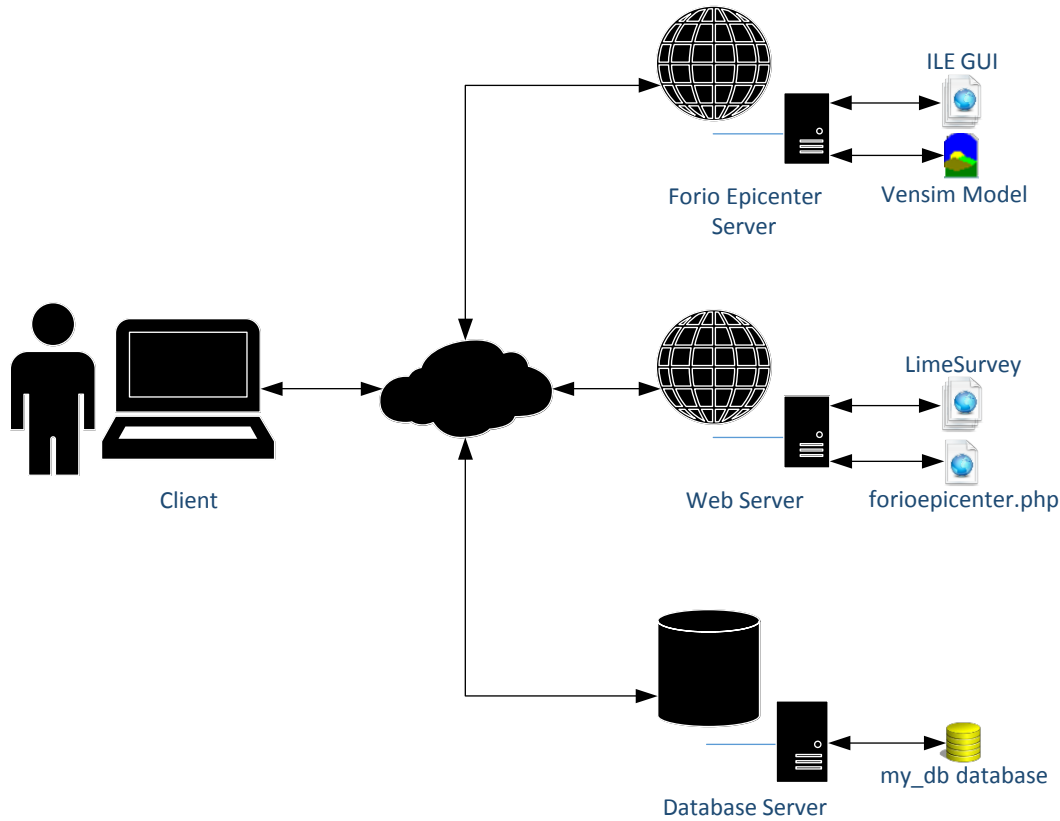


Figure 5: Dashboard tab



## 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 the decisions they take and their results. To accomplish such task, Epicenter uses Node.js<sup>11</sup> 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 framework that could be used by everyone.



**Figure 6: ILE system framework**

To log users' interactions, we have developed JavaScript snippets<sup>12</sup> and added them to all decision control elements (policy options in the control panel of the client-side) and charts (charts in the dashboard of the client-side) available on the GUI. These JavaScript snippets communicate with a PHP file called forioepicenter.php.<sup>13</sup> We developed forioepicenter.php

<sup>11</sup> <https://nodejs.org>

<sup>12</sup> The snippets are available inside HTML of the ILE. It could be shown by viewing the page source using any web browser. Furthermore, we have made these snippets generic, and marked them by HTML comment “<!--begin ” and “<!--end ”, to be easily copied to any other ILE.

<sup>13</sup> forioepicenter.php can be deployed to any server/web hotel supporting PHP, available at: <https://forio.com/app/ahmedg/eparticipation/helper/forioepicenter.php>

to save the values sent by the GUI to MySQL database.<sup>14</sup> Finally, a survey tool was needed to deploy our pre- and post-test questionnaires, we opted for Limesurvey.<sup>15</sup> The ILE system framework is shown in Figure 6.

## Procedure

We have conducted an expert opinion poll and 2 different experiments with users. For the expert opinion poll, we have invited 7 experts in the field of disability employment in Norway. The expert opinion poll session was totally online via video conference. Though, only 3 of the experts could attend the session on the 2<sup>nd</sup> of September 2015. Their ages covered age groups of 35-44, 45-54, and 55-64. While their educational degrees ranged between bachelors and doctorate degrees, and their years of experience in the field of disability employment in Norway ranged between 7 and 20 years. The poll began by a presentation of around 30 minutes about the model, the ILE, and the experimental setting. The video conference session was ended by sending a URL invitation to each expert to connect to our experiment server via her/his web-browser. The video conference stopped during the experiment session that was supposed to take no more than 40 minutes to answer the questionnaires and use the ILE. The video conference was resumed afterwards to discuss the experts' experience and hear their comments on the text of the experimental questionnaires and the ILE interface, which were used afterwards to enhance these texts. It was discovered that not all of them had finished the whole experiment; so based on where each of them has stopped in the experiment, we followed up by customising a continuation session and sending her/him its URL. The last response collected was on the 7<sup>th</sup> of September 2015. In addition to their answers to the experimental questionnaires, they have answered an additional set of questions about the usefulness and expected use of the ILE. Table 1 summarises the dates and number of participants in our expert opinion poll and experiments.

Our 1<sup>st</sup> experiment with users was conducted on the 13<sup>th</sup> of July 2015 in Cairo University, Egypt. This experiment was conducted with the available students attending a course provided by EDUEgypt programme<sup>16</sup> at the university by that time, whom were 14 students. By the end of the experiment, we could extract 11 completed and useful surveys. Table 2 shows demographic data of these participants. Subjects of our 2<sup>nd</sup> experiment were volunteer students from the University of Agder (UiA), Norway, on the 8<sup>th</sup> of September 2015. A couple of weeks earlier we started spreading the invitation for a gameplay session with free pizza in Grimstad campus of the university. At the day of the experiment, 17 students showed up. Some faced technical troubles with the experimentation system, and by the end we could extract 10 useful finished surveys. Their demographic data are presented in Table 3.

In all cases, the experimental session took around 1 hour. For 20 minutes, we gave a presentation to introduce the participants to the topic and the ILE. The presentation

---

<sup>14</sup> MySQL database tables needed by forioepicenter.php, can be reproduced in any MySQL using my\_db.sql, available at: [https://forio.com/app/ahmedg/eparticipation/helper/my\\_db.sql](https://forio.com/app/ahmedg/eparticipation/helper/my_db.sql)

<sup>15</sup> <https://www.limesurvey.org>

<sup>16</sup> The Education Development for the Universities of Egypt (EDUEgypt) is one of the programmes of the *Information Technology Institute (ITI)*, Egypt (<http://www.iti.gov.eg/>).

included the terms which 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 the case of EDUEgypt experiment, the text of the presentation and the ILE referring to the Norwegian case was changed to a general case to decrease any probable confusion; also participants were informed about an honorarium of 150 Egyptian pounds for the 3 highest-performing participants. In the case of UiA, 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 by the pre-test questionnaire that was consisted of 13 Likert 5-point scale items (strongly disagree, disagree, neutral, agree, or strongly agree).<sup>17</sup> These Likert items constitute multiple-item scale unidimensional construct as described by Brodersen and Thornton (2011), designed to test the participants' knowledge about the system's causalities and possible policy options.

To build the statements of these Likert items, we have enumerated all model variables affecting the disabled people employment rate. Possible changes in the values of these variables (for example: increase in, decrease in) were listed with different combination of possible resulted changes on the disabled people employment rate (for example: increase after short delay, no effect, immediate decrease, etc.). These combinations constituted the statements of the Likert items. These statements were ordered according to their importance based on our knowledge of the system, what we wanted to show and test, and how much they are clear while using the ILE. Further, to suit the experiment duration, 13 of these statements were selected, keeping a balance between reversed and non-reversed statements, and mostly following the recommendations stated by Weijters and Baumgartner (2012). Finally we polished the wording of the final statements, for example, participants were asked to report their level of agreement or disagreement with this statement: "Allowing disabled people to combine both salary and welfare benefits in the same time, will immediately increase employed disabled people".<sup>18</sup>

**Table 1: Accepted surveys**

Survey	Date	Participants	
Experts (online), Norway	02.09.2015/07.09.2015	7 invited	3 attended
EDUEgypt, Cairo University, Egypt	13.07.2015	14 attended	11 accepted surveys
University of Agder, Norway	09.09.2015	16 showed-up	10 accepted surveys

**Table 2: Properties of participants whose surveys were accepted – EDUEgypt experiment**

Property	Value	%
Age Group	18-24	91%
	25-34	9%
Gender	Female	64%

<sup>17</sup> Only in the case of EDUEgypt, the Likert items consisted of 15 items; they were decreased afterwards to shorten questionnaire time.

<sup>18</sup> All the Likert statements are available at: <https://forio.com/app/ahmedg/eparticipation/helper/questionnaires.txt>

Property	Value	%
	Male	36%
Field of Study/Work	Pharmacy	100%
Knowledge of Math Modelling	Yes	0%
Knowledge of System Thinking/Dynamics	Yes	0%

**Table 3: Properties of participants whose surveys were accepted – UiA experiment**

Property	Value	%
Age Group	18-24	70%
	25-34	30%
Gender	Male	80%
	Female	20%
Field of Study/Work	Information and Communication Technology (ICT)	30%
	Computer Engineering	20%
	Economics	20%
	Engineering	10%
	Mechatronics	10%
	Renewable Energy	10%
	Mechatronics	10%
Knowledge of Math Modelling	Yes	40%
Knowledge of System Thinking/Dynamics	Yes	30%

The Pre-test questionnaire was supposed to take no more than 7 minutes; nevertheless it was left to the participants to take as much time as they needed. The participants were informed that they could ask us for help all the time; however we abstained from providing any help that could lead to biases in participant answers to the questionnaire.

For a participant, the intervention using the ILE or the gameplay started as she/he ended the Pre-test questionnaire, without the option of going back to the Pre-test. The gameplay was limited to 25 minutes. During these 25 minutes, participants were instructed to imagine themselves as decision-makers responsible for creating a better situation for disabled people, and ideally save public spending. After finishing the gameplay, all participants were automatically directed to the post-test questionnaire, without the option of going back to the gameplay session. This way we were sure that all participants had not used the ILE for more than the designated duration.

The Post-test questionnaire contained exactly the same Likert items used in the Pre-test questionnaire. However once finished answering how she/he thinks now about each statement after using the ILE, the participant was asked to think back and report how much she/he agreed or disagreed with the same statement in the beginning of the session, based on her/his new understanding. This is called Retrospective Pre-test or Then-test (Howard 1980). It is very common that participants change their understanding between Pre-test and Post-test (Rockwell and Kohn 1989). 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 (Rockwell and Kohn 1989).

### **$\alpha$ , $\beta$ , and $\gamma$ Change**

Different SD literature presented methods to measure changes in mental models, or to compare them, for example (Markiczy and Goldberg 1995; Doyle, Radzicki, and Trees 2008; Schaffernicht and Groesser 2009; Schaffernicht and Groesser 2011; Groesser and Schaffernicht 2012). These methods either need a human rater, or require the test subjects

to have prior knowledge about certain knowledge elicitation tools, for example Causal Loop Diagrams (CLD). We were interested in a method free from these requirements. Human raters might cause experimenter bias (Campbell and Stanley 1963), while there was no guarantee that our test subjects would have enough knowledge about any knowledge elicitation tools.

Golembiewski *et al.* (1976) distinguished among 3 different types of attitude change as a result of an intervention, namely  $\alpha$ ,  $\beta$ , and  $\gamma$ .  $\alpha$  change refers to an absolute quantitative change (Riordan *et al.* 2001). For example, a person might “agree” that “spending on an awareness campaign to reach all disabled people potential employers to convince them to recruit disabled people will immediately cause a decrease in disabled people unemployment rates”. After the intervention, this person’s level of agreement about the same statement increases to “strongly agree”. This is a real change in her/his opinion on a fixed measurement scale, or  $\alpha$  change.

$\beta$  change refers to a measurement scale intervals recalibration, i.e. a redefining in the measurement standards. For example, a person has certain understanding of the values of different agreement levels (strongly disagree, disagree ... etc.) regarding a claim like “spending on an awareness campaign to reach all disabled people potential employers to convince them to recruit disabled people will immediately cause a decrease in disabled people unemployment rates”. Based on this understanding, this person indicates that she/he “strongly agrees” with that claim. After the intervention, this person finds out that what she/he used to interpret as “strongly agree” means just “agree”, consequently his answer to such a claim would change, although this does not reflect any change in her/his opinion. This is a change in the measurement continuum, or a change in the measurement standard /scale, or  $\beta$  change.

$\gamma$  change refers to a conceptual change, i.e. a redefinition of the measurement construct (Golembiewski, Billingsley, and Yeager 1976). For example, a person might have no idea that “spending on an awareness campaign to reach all disabled people potential employers to convince them to recruit disabled people will immediately cause a decrease in disabled people unemployment rates” or about “awareness campaigns targeting disabled people potential employers” at all. After the intervention, this person is provided with an understanding of the “awareness campaigns” and what they can do, i.e. a new conceptual frame of reference, which causes a meaningful answer based on this new understanding, or  $\gamma$  change.

Many methods to assess  $\alpha$ ,  $\beta$ , and  $\gamma$  changes appeared since 1976, including the method suggested by Golembiewski and his colleagues (1976). According to a comprehensive literature review conducted by Riordan *et al.* (2001), there are 5 major methods to detect  $\alpha$ ,  $\beta$ , and  $\gamma$  changes:

1. Ahmavaara's technique (Golembiewski, Billingsley, and Yeager 1976)
2. Actual-ideal difference measures (Zmud and Armenakis 1978)
3. Retrospective accounts (Terborg, Howard, and Maxwell 1980)
4. Confirmatory factor analysis (Schmitt 1982)
5. Latent growth modelling (Chan 1998)

Our expectations about the number of participants taking part in our experiment were 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 does not require a large sample, in addition to that it can test for  $\alpha$ ,  $\beta$ , and  $\gamma$  change independently (Riordan *et al.* 2001). Furthermore, it's 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 (2003).

### Retrospective Accounts

Since Post-test and Then-test are answered based on the same understanding/perception as mentioned above, the Retrospective accounts method detects  $\alpha$  change by detecting the change between them. Furthermore, Pre-test and Then-test are basically measuring the same thing based on either 2 different understandings/perceptions or 2 differently calibrated measurement scales,  $\gamma$  and  $\beta$  change are detected by detecting the change between Pre-test and Then-test (Terborg, Howard, and Maxwell 1980).

The Retrospective accounts method supports analysis on both group and individual levels. However 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. Occasionally certain individual change could be covered by detecting just group changes (Birkenbach 1986; Brodersen and Thornton 2011). Furthermore, "*a large amount of change exhibited by only a few individuals may be taken as evidence that the intervention had a group effect*" (Brodersen and Thornton 2011).

To apply the Retrospective accounts method to our collected data, we have followed the practice of Birkenbach (1986) in general. Nevertheless, we have opted for following Brodersen and Thornton (2011) in detecting  $\gamma$  change first, then remove the participants showing  $\gamma$  change from the process of detecting  $\alpha$  and  $\beta$ . According to Porras and Singh (1986) when  $\gamma$  change is detected, the detection of  $\alpha$  or  $\beta$  becomes problematic.

Answers to questionnaire items from Pre-, Post-, and Then-tests of each participant were used as raw data/basic data points (Porras and Singh 1986). So for every participant, we have compiled 3 paired samples **Pre**, **Post**, and **Then**. The first step is to try detecting  $\gamma$  change per participant. Terborg and his colleagues (1980) suggested 2 methods:

1. *Using Correlation:*

For every participant, correlations between the following pairs are calculated:

- **Pre and Then** ( $r_{\text{Pre Then}}$ )
- **Post and Pre** ( $r_{\text{Post Pre}}$ )
- **Post and Then** ( $r_{\text{Post Then}}$ )

To test for differences between the pairs  $r_{\text{Post Then}}$  &  $r_{\text{Pre Then}}$  and  $r_{\text{Post Then}}$  &  $r_{\text{Post Pre}}$ , Williams's test<sup>19</sup> to compare correlations of 2 paired/dependant samples is used to calculate  $t_{(r_{\text{Post Then}})(r_{\text{Pre Then}})}$  and  $t_{(r_{\text{Post Then}})(r_{\text{Post Pre}})}$  (Brodersen and Thornton 2011).  $\gamma$  change exists if the following 2 conditions are met:

- a.  $r_{\text{Post Then}}$  is substantially greater than  $r_{\text{Pre Then}}$
- b.  $r_{\text{Post Then}}$  is substantially greater than  $r_{\text{Post Pre}}$

---

<sup>19</sup> To apply Williams's Test to test the difference between 2 dependent correlations sharing 1 variable/2 "paired" correlations, we used the R (The R project for statistical computing software environment <https://www.r-project.org>) command: `r.test {package: psych}` (<http://www.personality-project.org/r/html/r.test.html>).

## 2. Using Standard Deviation:

**Pre**, **Post**, and **Then** Standard Deviations are calculated for every participant, yielding  $s_{\text{Pre}}$ ,  $s_{\text{Post}}$ , and  $s_{\text{Then}}$  respectively. Morgan-Pitman test<sup>20</sup> 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  $t_{(s_{\text{Pre}})(s_{\text{Post}})}$ .  $\gamma$  change exists if the following conditions are met:

- a.  $s_{\text{Post}}$  is not different from  $s_{\text{Then}}$
- b.  $s_{\text{Post}}$  is different from  $s_{\text{Pre}}$
- c.  $s_{\text{Then}}$  is different from  $s_{\text{Pre}}$

The highest level of  $\gamma$  change happens when both correlation and standard deviation methods to detect  $\gamma$  change occur concurrently (Terborg, Howard, and Maxwell 1980). If a participant does not show any signs of  $\gamma$  change, we start detecting  $\beta$  or  $\alpha$  change.

To test for  $\beta$  or  $\alpha$  change, mean values of **Pre**, **Post**, and **Then** are calculated for every participant, yielding  $\bar{x}_{\text{Pre}}$ ,  $\bar{x}_{\text{Post}}$ , and  $\bar{x}_{\text{Then}}$  respectively. Student's t-Test<sup>21</sup> to compare means of 2 paired/dependant samples is used to calculate  $t_{(\text{Then})(\text{Pre})}$  and  $t_{(\text{Then})(\text{Post})}$ . If  $t_{(\text{Then})(\text{Post})}$  is greater than  $t_{(\text{Then})(\text{Pre})}$ , descriptively speaking there is more evidence of  $\alpha$  change than  $\beta$  change, and vice versa (Terborg, Howard, and Maxwell 1980). Following the practice of Birkenbach (1986), we have focused only on the size to compare  $t_{(\text{Then})(\text{Pre})}$  to  $t_{(\text{Then})(\text{Post})}$ .

Terborg and his colleagues (1980) emphasised on that t-statistics on the individual level analysis should generally 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 (Terborg, Howard, and Maxwell 1980).

## Results and Discussion

### Results of the Expert Opinion Poll

Our ILE was found easy-to-use by 67% of our poll of experts. The same percentage thinks that this ILE could be used in decision-making; its target audience could be decision-makers in general, and particularly decision-makers within disabled people organisations. All the experts of our expert opinion poll agreed on that the ILE could support targeted discussions among citizens, organisations, policymakers about the employment problem of disabled people. Also, using the ILE to raise awareness and in

---

<sup>20</sup> To apply Morgan-Pitman test to test for equal variance of 2 dependent samples, we used R command: `var.test {package: PairedData}` (<http://artax.karlin.mff.cuni.cz/r-help/library/PairedData/html/var.test.html>).

<sup>21</sup> To apply Student's t-Test to compare means of 2 paired samples, we used R command: `t.test {package: stats}` (<https://stat.ethz.ch/R-manual/R-patched/library/stats/html/t.test.html>).

advocacy campaigns were suggested by the experts. 67% of the experts agreed that the ILE could be used to reconnect ordinary people with disabled people employment problem, and in that regard, they suggested using this ILE and its results in seminars, public articles, election campaigns, and TV programmes, both on central and local levels. In conclusion, 67% of the experts agreed that the ILE is useful in general.

### Results of the EDUEgypt Experiment

Participant P5 was removed from the analysis because of showing no variance in her **Post** sample ( $s_{\text{Post}}^2 = 0$ ). As mentioned above, the analysis starts by detecting  $\gamma$  change. This is done via correlation and standard deviation comparisons. The left half of Table 4 shows the needed correlation values in addition to the t-statistics calculated to compare them. The t-statistic columns at the left half of table prove that  $r_{\text{Post Then}}$  is substantially greater than  $r_{\text{Post Pre}}$  and  $r_{\text{Pre Then}}$  for participants P1, P10, and P11, consequently showing  $\gamma$  change. The t-statistics columns on the right half of the same table cannot at all prove that  $s_{\text{Post}}$  is not different from  $s_{\text{Then}}$ , while both are different from  $s_{\text{Pre}}$  for any participant, and consequently no  $\gamma$  change was detected based on standard deviation. After ignoring participants showing  $\gamma$  change, from  $\beta$  and  $\alpha$  change detection procedure, Table 5 shows that P2, P3, P4, P6, and P7 have smaller values of  $t_{(\text{Then})(\text{Post})}$  compared to  $t_{(\text{Then})(\text{Pre})}$  denoting  $\beta$  change for these participants. Accordingly participants P8, and P9 have exhibited  $\alpha$  change. Table 6 show the overall  $\alpha$ ,  $\beta$ , and  $\gamma$  changes detected for all participants in comparison to their answers about mathematical modelling and system dynamics knowledge. Clearly, there is no association between these variables and the detected  $\alpha$ ,  $\beta$ , or  $\gamma$  changes.

### Results of the UiA Experiment

The t-statistics columns at the left half of Table 7 prove that  $r_{\text{Post Then}}$  is substantially greater than  $r_{\text{Post Pre}}$  and  $r_{\text{Pre Then}}$  for participants P1, P5, P7, and P8, consequently showing  $\gamma$  change. The t-statistics columns on the right half of the same table cannot at all prove that  $s_{\text{Post}}$  is not different from  $s_{\text{Then}}$ , while both are different from  $s_{\text{Pre}}$  for any participant, and consequently no  $\gamma$  change was detected based on standard deviation. After ignoring participants showing  $\gamma$  change, from  $\beta$  and  $\alpha$  change detection procedure, Table 8 shows that P3, P9, and P10 have smaller values of  $t_{(\text{Then})(\text{Post})}$  compared to  $t_{(\text{Then})(\text{Pre})}$  denoting  $\beta$  change for these participants. Therefore, participants P2, P4, and P6 have exhibited  $\alpha$  change. Table 9 show the overall  $\alpha$ ,  $\beta$ , and  $\gamma$  changes detected for all participants in comparison to their answers about mathematical modelling and system dynamics knowledge. We could find weak associations between the mathematical modelling knowledge on one side, and the detected  $\alpha$  and  $\beta$  changes on the other side, with phi coefficient of 0.36 and -0.53 respectively (Simon 2016).



## General Discussion

The expert opinion poll provided us with a qualitative measure to assess the ILE and consequently the model, whereas the 2 experiments conducted with users from 2 different backgrounds and different geographical areas provided us with a quantitative measure. Our expert opinion poll results show that 67% of the experts think that the ILE achieves its intended goals. Meanwhile, 33% of the participants who were included the analysis of both experiments have shown a change in their understanding and perceptions of the system's causal relationships and policy options. Moreover, 38% have redefined/recalibrated the standards they use to assess or evaluate these relationships and policy options. In total, 71% of the participants have redefined certain knowledge as a result of using the ILE, achieving the ILE's intended goals.

From an internal validity (Campbell and Stanley 1963) point of view, to minimise testing validity threat, in all cases we have kept the questionnaires as merely Likert-scale items, and emphasised to participants that there is no right or wrong answer, they need to report what they thought/believed. Furthermore, we made sure that all participants have fully understood questionnaire items since the pre-test, to account for any misunderstandings 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. Nevertheless, we have to admit that the research suffered from selection validity threat due to the availability of participants as previously mentioned.

Furthermore, from the external validity (Bracht and Glass 1968) perspective, although participants of both experiments were few and limited to university students, they were from 2 different backgrounds and different geographical areas. Yet, other experiments with different samples are necessary. Furthermore, longer periods between pre-test, treatment, and 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 too.

**Table 4:  $\gamma$  change detection – EDUEgypt Experiment**

	$r_{\text{Post Then}}$	$r_{\text{Pre Post}}$	$r_{\text{Pre Then}}$	$t_{(r_{\text{Post Then}})(r_{\text{Pre Post}})}$	$t_{(r_{\text{Post Then}})(r_{\text{Pre Then}})}$	$\gamma$ change detected	$S_{\text{Post}}$	$S_{\text{Then}}$	$S_{\text{Pre}}$	$t_{(S_{\text{Post}})(S_{\text{Then}})}$	$t_{(S_{\text{Pre}})(S_{\text{Post}})}$	$t_{(S_{\text{Pre}})(S_{\text{Then}})}$	$\gamma$ change detected
P1	0.81 <sup>***</sup>	0.46 <sup>*</sup>	0.57 <sup>**</sup>	2.21 <sup>**</sup>	1.40 <sup>*</sup>	Yes	1.11	1.29	1.26	-0.92	-0.52	-0.09	--
P2	0.51 <sup>**</sup>	0.88 <sup>***</sup>	0.44 <sup>*</sup>	-2.39 <sup>**</sup>	0.52	--	1.40	1.31	1.49	0.29	-0.43	0.51	--
P3	0.18	0.17	0.21	0.04	-0.07	--	0.46	0.74	0.62	-1.85 <sup>*</sup>	-1.11	-0.69	--
P4	-0.19	-0.23	-0.35	0.08	0.38	--	0.83	0.90	1.06	-0.31	-0.92	0.63	--
P5 <sup>X</sup>	--	--	-0.08	--	--	--	0.00	0.35	1.01	--	--	4.59 <sup>***</sup>	--
P6	0.28	0.02	0.22	0.74	0.16	--	0.99	0.41	1.26	3.71 <sup>***</sup>	-0.89	5.04 <sup>***</sup>	--
P7	0.22	0.20	0.49 <sup>*</sup>	0.08	-0.85	--	0.52	0.51	0.52	0.07	0.00	0.08	--
P8	0.47 <sup>*</sup>	0.27	0.49 <sup>*</sup>	0.76	-0.08	--	0.41	0.52	0.51	-0.91	-0.76	-0.08	--
P9	0.31	0.05	0.31	0.83	0.01	--	1.36	0.59	1.40	3.50 <sup>***</sup>	-0.13	3.68 <sup>***</sup>	--
P10	0.46 <sup>*</sup>	0.88 <sup>***</sup>	0.76 <sup>***</sup>	-5.74 <sup>***</sup>	-4.29 <sup>***</sup>	Yes	0.52	1.59	1.05	-5.60 <sup>***</sup>	-5.85 <sup>***</sup>	-2.38 <sup>**</sup>	--
P11	0.96 <sup>***</sup>	0.64 <sup>***</sup>	0.62 <sup>***</sup>	3.80 <sup>***</sup>	4.15 <sup>***</sup>	Yes	0.59	0.38	0.92	5.68 <sup>***</sup>	-2.10 <sup>*</sup>	4.61 <sup>***</sup>	--

\* P < 0.10 \*\* P < 0.05 \*\*\* P < 0.01

<sup>X</sup> Participant removed because of showing no variance in Pre, Post, and/or Then

**Table 5:  $\alpha$  and  $\beta$  change detection – EDUEgypt Experiment**

	$\bar{x}_{\text{Post}}$	$\bar{x}_{\text{Then}}$	$\bar{x}_{\text{Pre}}$	$t_{(\text{Then})(\text{Post})}$	$t_{(\text{Then})(\text{Pre})}$	<b><math>\beta</math> change detected</b>	<b><math>\alpha</math> change detected</b>
P1 <sup>G</sup>	3.67	3.67	3.80	0.00	0.43	--	--
P2	3.40	3.00	3.73	1.15	1.91*	Yes	--
P3	4.07	4.13	4.33	-0.32	0.90	Yes	--
P4	3.60	3.67	3.87	-0.19	0.48	Yes	--
P5 <sup>X</sup>	4.00	4.87	3.80	-9.54***	-3.76**	--	Yes
P6	4.13	4.20	3.80	-0.27	-1.25	Yes	--
P7	4.47	4.60	4.47	-0.81	-1.00	Yes	--
P8	4.20	4.53	4.40	-2.65**	-1.00	--	Yes
P9	3.53	3.93	3.60	-1.19	-0.96	--	Yes
P10 <sup>G</sup>	3.87	3.67	4.33	0.54	2.47**	--	--
P11 <sup>G</sup>	3.93	4.00	3.87	-1.00	-0.69	--	--

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

<sup>X</sup> Participant removed because of showing no variance in Pre, Post, and/or Then

<sup>G</sup>  $\gamma$  change detected

**Table 6:  $\alpha$ ,  $\beta$ , and  $\gamma$  change results vs participants' properties – EDUEgypt Experiment**

	<b>Knowledge of Math Modelling</b>	<b>Knowledge of System Thinking/Dynamics</b>	<b><math>\gamma</math> change detected</b>	<b>B change detected</b>	<b><math>\alpha</math> change detected</b>
P1	--	--	Yes	--	--
P2	--	--	--	Yes	--
P3	--	--	--	Yes	--
P4	--	--	--	Yes	--
P5 <sup>x</sup>	--	--	--	--	--
P6	--	--	--	Yes	--
P7	--	--	--	Yes	--
P8	--	--	--	--	Yes
P9	--	--	--	--	Yes
P10	--	--	Yes	--	--
P11	--	--	Yes	--	--

<sup>x</sup> Participant removed because of showing no variance in Pre, Post, and/or Then

**Table 7:  $\gamma$  change detection – UiA Experiment**

	$r_{\text{Post Then}}$	$r_{\text{Pre Post}}$	$r_{\text{Pre Then}}$	$t_{(r_{\text{Post Then}})(r_{\text{Pre Post}})}$	$t_{(r_{\text{Post Then}})(r_{\text{Pre Then}})}$	$\gamma$ change detected	$s_{\text{Post}}$	$s_{\text{Then}}$	$s_{\text{Pre}}$	$t_{(s_{\text{Post}})(s_{\text{Then}})}$	$t_{(s_{\text{Pre}})(s_{\text{Post}})}$	$t_{(s_{\text{Pre}})(s_{\text{Then}})}$	$\gamma$ change detected
P1	0.95 <sup>***</sup>	0.43	0.46	4.31 <sup>***</sup>	3.92 <sup>***</sup>	Yes	0.55	0.90	0.96	-5.41 <sup>***</sup>	-2.12 <sup>*</sup>	0.25	--
P2	0.43	-0.11	0.45	1.94 <sup>**</sup>	-0.05	--	0.75	0.95	0.93	-0.89	-0.71	-0.11	--
P3	0.39	0.02	0.28	1.05	0.27	--	1.12	0.95	1.01	0.56	0.32	0.21	--
P4	0.74 <sup>***</sup>	0.59 <sup>*</sup>	0.57 <sup>**</sup>	0.77	0.90	--	0.85	1.39	1.13	-2.48 <sup>**</sup>	-1.15	-0.84	--
P5	0.80 <sup>***</sup>	0.14	-0.16	2.09 <sup>*</sup>	4.11 <sup>***</sup>	Yes	1.04	1.01	1.12	0.17	-0.24	0.34	--
P6	0.78 <sup>***</sup>	0.34	0.52 <sup>*</sup>	2.23 <sup>**</sup>	1.18	--	0.88	0.77	0.78	0.71	0.43	0.04	--
P7	0.95 <sup>***</sup>	0.78 <sup>***</sup>	0.83 <sup>***</sup>	2.94 <sup>***</sup>	1.94 <sup>**</sup>	Yes	1.20	1.19	0.99	0.05	1.03	-1.14	--
P8	1.00 <sup>***</sup>	0.60 <sup>**</sup>	0.60 <sup>**</sup>	8.66 <sup>***</sup>	8.66 <sup>***</sup>	Yes	0.80	0.80	0.99	n/a	-0.87	0.87	--
P9	0.55 <sup>**</sup>	0.63 <sup>**</sup>	0.53 <sup>**</sup>	-0.36	0.06	--	0.73	0.85	0.91	-0.66	-0.99	0.26	--
P10	0.69 <sup>***</sup>	0.55 <sup>**</sup>	0.75 <sup>**</sup>	0.88	-0.32	--	1.15	1.14	1.04	0.04	0.40	-0.46	--

\* P < 0.10 \*\* P < 0.05 \*\*\* P < 0.01

**Table 8:  $\alpha$  and  $\beta$  change detection – UiA Experiment**

	$\bar{x}_{\text{Post}}$	$\bar{x}_{\text{Then}}$	$\bar{x}_{\text{Pre}}$	$t_{(\text{Then})(\text{Post})}$	$t_{(\text{Then})(\text{Pre})}$	<b><math>\beta</math> change detected</b>	<b><math>\alpha</math> change detected</b>
P1 <sup>G</sup>	3.85	3.85	3.38	0.00	-1.72	--	--
P2	3.31	3.08	3.23	0.90	0.56	--	Yes
P3	3.08	3.08	3.77	0.00	2.11*	Yes	--
P4	3.69	3.38	3.46	1.17	0.23	--	Yes
P5 <sup>G</sup>	3.38	3.77	3.62	-2.13**	-0.34	--	--
P6	3.46	3.62	3.54	-1.00	-0.37	--	Yes
P7 <sup>G</sup>	3.54	3.38	3.85	1.48	2.52**	--	--
P8 <sup>G</sup>	3.85	3.85	3.85	n/a	0.00	--	--
P9	3.23	3.31	3.00	-0.37	-1.30	Yes	--
P10	3.00	3.15	3.62	-0.62	2.14**	Yes	--

\* P < 0.10 \*\* P < 0.05 \*\*\* P < 0.01

<sup>G</sup>  $\gamma$  change detected

**Table 9:  $\alpha$ ,  $\beta$ , and  $\gamma$  change results vs participants' properties – UiA Experiment**

	<b>Knowledge of Math Modelling</b>	<b>Knowledge of System Thinking/Dynamics</b>	<b><math>\gamma</math> change detected</b>	<b><math>\beta</math> change detected</b>	<b><math>\alpha</math> change detected</b>
P1	Yes	Yes	Yes	--	--
P2	Yes	--	--	--	Yes
P3	--	Yes	--	Yes	--
P4	--	Yes	--	--	Yes
P5	--	--	Yes	--	--
P6	Yes	--	--	--	Yes
P7	Yes	--	Yes	--	--
P8	--	--	Yes	--	--
P9	--	--	--	Yes	--
P10	--	--	--	Yes	--

## Conclusion

In Norway, the unemployment rate of disabled people is unchanged since 2000, making the country to perform less compared to the neighbouring countries, despite the repeated governmental promises. These low rates prevail even under the protection of anti-discrimination law and high governmental spending on disabled people employment efforts compared to any other country. To increase disabled people employment probability, the Norwegian government has introduced different vocational rehabilitation programmes. Unfortunately, these programmes suffer from many problems. On the contrary Telenor's small-scale Open Mind programme is a very successful rehabilitation programme.

"Disabled People from Welfare to Jobs" is a SD model that was inspired by the idea of operationalising Open Mind-like programme country-wide. We have developed an ILE to present this model. The model is intended to support more targeted discussions among individual citizens, disability organisations, and policy makers, and to contribute in linking ordinary people with policymaking and rendering the decision-making processes easier to comprehend. This requires the model to be able to change how its users think and take decisions as well as to change their understanding, perceptions, and accordingly attitudes. Our goal was to test whether this model is really capable of doing what it is intended to or not. For that, we have conducted expert opinion poll, and 2 experiments with the ILE.

67% of the expert of our expert opinion poll thinks that the ILE achieves its intended goals. Furthermore, we have conducted  $\alpha$ ,  $\beta$ , and  $\gamma$  change analysis on the results of the 2 experiments, on the individual level. 33% of the participants who were included the analysis of both experiments has shown a change in their understanding and perceptions of the system's causalities and policy options. Meanwhile, 38% have redefined/recalibrated the standards they use to assess or evaluate these relationships and policy options. In total, 71% of the participants have redefined certain knowledge as a result of using the ILE, or have their mental models changed, achieving the ILE's intended goals.

In this paper we have also provided a methodological contribution. We have developed a generic reusable ILE framework, and provided instructions on how it could be used by others in creating their ILEs. Furthermore, we have adapted the  $\alpha$ ,  $\beta$ , and  $\gamma$  change typology and the retrospective accounts method to test the effect of using an ILE on its users. We have also introduced our suggested approach to create the questionnaires needed to apply the  $\alpha$ ,  $\beta$ , and  $\gamma$  change and the retrospective accounts method in testing an ILE effect on its users, as well as our suggested steps and statistical tests needed in conducting the statistical analysis for the retrospective accounts method.

As a final point, applying  $\alpha$ ,  $\beta$ , and  $\gamma$  change analysis to test the effect of using SD based ILE was easy and straight forward. However, more experimentation with larger samples, ideally including control groups, to test for group changes in addition to individual changes, over longer time spans, and longer questionnaire seems to be a very promising and highly recommended future research. Furthermore, comparing the  $\alpha$ ,  $\beta$ , and  $\gamma$  change results with results from other mental model change measurement methods more common among SD practitioners is a very important validation requirement for the method in the SD field.



## Acknowledgement

We thank Dr. Heba Saleh, chairman of the Information Technology Institute (ITI), Egypt, for her support in facilitating the experiment in Egypt. Likewise, we thank all EDUEgypt students who participated in this experiment. We also thank Kamil Gryga and Dr. Morten Goodwin for their support in facilitating the experiment in Norway. Similarly, we thank all the students who volunteered to participate in this experiment.

## References

- Abdelgawad, A., M. Snaprud, J. Krogstie, and R. Brynn. 2012. 'Disabled People from Welfare to Jobs: A Decision Support Tool'. In *Modelling Symposium (AMS), 2012 Sixth Asia*, 132–37. Bali, Indonesia: IEEE Computer Society Conference Publishing Services'. doi:10.1109/AMS.2012.24.
- Bengtsson, Steen. 2011. 'Denmark – ANED Country Profile'. Academic Network of European Disability experts (ANED). <http://www.disability-europe.net/content/aned/media/DK%20-%20country%20page%20update%20-%20final.doc>.
- Birkenbach, Xavier C. 1986. 'Self-Report Evaluations of Training Effectiveness: Measuring Alpha, Beta, and Gamma Change'. *South African Journal of Psychology* 16 (1): 1–7. doi:10.1177/008124638601600101.
- Bracht, Glenn H., and Gene V. Glass. 1968. 'The External Validity of Experiments'. *American Educational Research Journal* 5 (4): 437–74. doi:10.2307/1161993.
- Brodersen, D. Apryl, and George C. Thornton. 2011. 'An Investigation of Alpha, Beta, and Gamma Change in Developmental Assessment Center Participants'. *Performance Improvement Quarterly* 24 (2): 25–48. doi:10.1002/piq.20109.
- Campbell, D. T., and J. C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. 1 edition. Belmont, CA: Wadsworth Publishing.
- Chan, David. 1998. '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* 1 (4): 421–83. doi:10.1177/109442819814004.
- Doyle, James K., Michael J. Radzicki, and W. Scott Trees. 2008. 'Measuring Change in Mental Models of Complex Dynamic Systems'. In *Complex Decision Making*, edited by H. Qudrat-Ullah, J. M. Spector, and P. I. Davidsen, 269–94. Understanding Complex Systems. Springer Berlin Heidelberg. [http://link.springer.com/chapter/10.1007/978-3-540-73665-3\\_14](http://link.springer.com/chapter/10.1007/978-3-540-73665-3_14).
- Friedman, Sheldon, Steven Cavaleri, and Michael Raphael. 2003. '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. <http://www.systemdynamics.org/conferences/2003/proceed/PAPERS/918.pdf>.
- Galis, Vasilis. 2012. 'Sweden – ANED Country Profile'. Academic Network of European Disability experts (ANED). <http://www.disability-europe.net/content/aned/media/SE%20-%20country%20webpage%20update%20-%20final.doc>.
- Golembiewski, Robert T., Keith Billingsley, and Samuel Yeager. 1976. 'Measuring Change and Persistence in Human Affairs: Types of Change Generated by OD Designs'. *The Journal of Applied Behavioral Science* 12 (2): 133–57. doi:10.1177/002188637601200201.

- Groesser, Stefan N., and Martin Schaffernicht. 2012. 'Mental Models of Dynamic Systems: Taking Stock and Looking Ahead'. *System Dynamics Review* 28 (1): 46–68. doi:10.1002/sdr.476.
- Howard, George S. 1980. 'Response-Shift Bias A Problem in Evaluating Interventions with Pre/Post Self-Reports'. *Evaluation Review* 4 (1): 93–106. doi:10.1177/0193841X8000400105.
- Markíczy, Livia, and Jeff Goldberg. 1995. 'A Method for Eliciting and Comparing Causal Maps'. *Journal of Management* 21 (2): 305–33. doi:10.1177/014920639502100207.
- OECD. 2003. *Transforming Disability into Ability*. OECD Publishing. [http://www.oecd.org/document/14/0,3343,en\\_2649\\_34747\\_35290126\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/14/0,3343,en_2649_34747_35290126_1_1_1_1,00.html).
- Porras, Jerry I., and Jitendra V. Singh. 1986. 'Alpha, Beta, and Gamma Change in Modelling-Based Organization Development'. *Journal of Organizational Behavior* 7 (1): 9–24. doi:10.1002/job.4030070104.
- Riordan, Christine M., Hettie A. Richardson, Bryan S. Schaffer, and Robert J. Vandenberg. 2001. 'Alpha, Beta, and Gamma Change: A Review of Past Research with Recommendations for New Directions'. In *Equivalence in Measurement*, edited by Linda L. Neider. IAP.
- Rockwell, S. Kay, and Harriet Kohn. 1989. 'Post-Then-Pre Evaluation'. *Journal of Extension* 27 (2). <http://www.joe.org/joe/1989summer/a5.php>.
- Røed, Knut, and Oddbjørn Raaum. 2006. 'Do Labour Market Programmes Speed up the Return to Work?' *Oxford Bulletin of Economics and Statistics* 68 (5): 541–68. doi:10.1111/j.1468-0084.2006.00177.x.
- Schaffernicht, Martin, and Stefan N. Groesser. 2009. 'What's in a Mental Model of a Dynamic System? Conceptual Structure and Model Comparison'. In *Proceedings of the 27th International Conference of the System Dynamics Society*. Albuquerque, New Mexico, USA. <http://www.systemdynamics.org/conferences/2009/proceed/papers/P1281.pdf>.
- . 2011. 'A Comprehensive Method for Comparing Mental Models of Dynamic Systems'. *European Journal of Operational Research* 210 (1): 57–67.
- Schmitt, Neal. 1982. 'The Use of Analysis of Covariance Structures to Assess Beta and Gamma Change'. *Multivariate Behavioral Research* 17 (3): 343–358.
- Simon, Steve. 2016. 'Stats: What Is a Phi Coefficient?' Accessed March 21. <http://www.pmean.com/definitions/phi.htm>.
- Skøien, Reidun, Karl-Gerhard Hem, and Geir Tyrmi. 2006. 'Evaluation of the Handicap-Programme at Telenor'. STF78 A06008. SINTEF Health Research. [http://www.telenor.com/en/resources/images/TelenorHCPReport\\_tcm28-35286.pdf](http://www.telenor.com/en/resources/images/TelenorHCPReport_tcm28-35286.pdf).
- Statistics Norway. 2016. 'Population Aged 15-66, Total, and Persons with Disabilities, by Sex and Labour Force Status'. *Statistics Norway*. <https://www.ssb.no/statistikkbanken/SelectVarVal/Define.asp?MainTable=Akutu02&KortNavnWeb=akutu&PLanguage=1&checked=true>.
- . 2016. 'Stable Employment Rate among Disabled People'. *Statistics Norway*. Accessed March 16. <http://www.ssb.no/en/arbeid-og-lonn/statistikker/akutu/aar/2015-09-02>.
- Sterman, John. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. McGraw Hill Higher Education.

- . 2014a. ‘Interactive Web-Based Simulations for Strategy and Sustainability: The MIT Sloan LearningEdge Management Flight Simulators, Part I’. *System Dynamics Review* 30 (1–2): 89–121. doi:10.1002/sdr.1513.
- . 2014b. ‘Interactive Web-Based Simulations for Strategy and Sustainability: The MIT Sloan LearningEdge Management Flight Simulators, Part II’. *System Dynamics Review* 30 (3): 206–31. doi:10.1002/sdr.1519.
- Telenor Group. 2012. ‘Telenor Group | For the Physically Challenged’. *Telenor Group*. March 7. <http://www.telenor.com/career/programme-for-work-integration/for-the-physically-challenged/>.
- Terborg, James R., George S. Howard, and Scott E. Maxwell. 1980. ‘Evaluating Planned Organizational Change: A Method for Assessing Alpha, Beta, and Gamma Change’. *The Academy of Management Review* 5 (1): 109–21. doi:10.2307/257811.
- Tøssebro, Jan. 2009. ‘Report on the Employment of Disabled People in European Countries - Norway’. Academic Network of European Disability experts (ANED). <http://www.disability-europe.net/content/pdf/NO - ANED 2009 Employment Report Final.pdf>.
- Weijters, Bert, and Hans Baumgartner. 2012. ‘Misresponse to Reversed and Negated Items in Surveys: A Review’. *Journal of Marketing Research* 49 (5): 737–47. doi:10.1509/jmr.11.0368.
- Westlie, Lars. 2008. ‘Norwegian Vocational Rehabilitation Programs: Improving Employability and Preventing Disability?’ Oslo, Norway: Oslo University, Department of Economics. [http://ideas.repec.org/p/hhs/osloec/2008\\_024.html](http://ideas.repec.org/p/hhs/osloec/2008_024.html).
- Zmud, Robert W., and Achilles A. Armenakis. 1978. ‘Understanding the Measurement of Change’. *Academy of Management Review* 3 (3): 661–69. doi:10.5465/AMR.1978.4305950.