International Journal of
Advanced Science Computing and Engineering

Journal Homepage www.ijasce.org/index.php/ijasce

Design and Validation of Structural Causal Model: A Focus on SENSE-EGRA Datasets

Gabriel Terna Ayem [a,*], Augustine Shey Nsang [a], Bernard Igoche Igoche [b], Garba Naankang [c]

[a] *Computer Science Department, School of Information Technology and Computing, American University of Nigeria, Yola, Nigeria*
[b] *Computer Science Department, School of Computing, University of Portsmouth, United Kingdom*
[c] *True-life Engineering Department, School of Aerospace, Transport, and Manufacturing, Cranfield University, United Kingdom*
*Corresponding author: *gabriel.ayem@aun.edu.ng*

***Abstract*—Designing and validating a causal model's correctness from a dataset whose background knowledge is obtained from a research process is not a common phenomenon. Studies have shown that in many critical areas, such as healthcare and education, researchers develop models from direct acyclic graphs without testing them. This phenomenon is worrisome and is bound to cast a dark shadow on the inference estimates that many arise from such models. In this study, we have designed a novel application-based SCM for the first time using the background knowledge gained from the American University of Nigeria (AUN), Yola, on the letter identification subtask of the Early Grade Reading Assessment (EGRA) program on the Strengthen Education in Northeast Nigeria (SENSE-EGRA) project dataset, which the USAID sponsored. We employed the conditional independence test (CIT) criteria for the model's correctness validation testing, and the results show a near-perfect SCM.**

***Keywords*—Causality; structural causal validation; model assumption; observational datasets; model testing.**

Manuscript received 10 Jan. 2024; revised 23 Feb. 2024; accepted 21 Mar. 2024. Date of publication 30 Apr. 2024.

I. INTRODUCTION

From time immemorial till date, human actions, processes, and indeed scientific explorations have been predicated on the premise of cause and effect. In the primordial era, the savaged and primitive man sought ways to articulate and uncover this phenomenon of cause and effects; and not having equipment, enough facts or the sine-quo-non to ascertain this phenomenon of knowing what actions (causes) that produces the effects mainly in incidences that were agonizing to him such as certain ebullitions of some sicknesses concomitant with mysterious deaths. Thus, the ability to know the right action to influence his environment or predict his future made man an idiosyncratic species from the rest of the animals. Thus, driving the savaged man from his initial state of higgledy-piggledy to embrace the practice of magic, astrology, and specific fetish ways to achieve the causation phenomenon to overcome his bewildered state. Gradually, as societies evolved and advanced, mankind himself advanced from his primitive and savage state to his current state of scientific and technological advancement. Thus, establishing his hegemony over every other species on earth, the same motives of trying to influence his environment and predict his future still stand.

Nonetheless, the methods of achieving it have evolved; as magic arts give way to scientific logic, astrology transforms into astronomy, and other technological innovations, such as computer predictions and simulations, become the modern genies that are aberrations from the fetish ways of predicting the future. Albeit in this current era, the science of trying to ascertain causality or causation in human processes and actions is still a daunting and nontrivial task, as the traditional scientific way of establishing this act is associated with the randomized controlled experiment or randomized controlled trial (RCT) method. This RCT method and idea are credited to Fisher [1].

Thus, this standard framework for causal discovery, known as RCT, always involves setting aside some (usually half) of the sampled population of the study and giving them treatment (an intervention) under the same conditions. In contrast, the second half of the study population is left untreated (not intervened on) or controlled under the same or similar conditions, to slay any possible confounding or lurking variable, which is often the factor that jeopardizes a proper juxtaposition of these two sampled populations in the RCT

experiments. As fascinating as this method of RCT is, some events and circumstances make this kind of experiment too expensive, infeasible, or even unethical to perform. A good instance is to conduct an RCT on a hypothesized query that seeks to uncover the health benefits, or otherwise, of smoking on a specific population. This is an unethical experiment to conduct under RCT, because it would involve setting half of the population under review to smoke (treated) and the other not to smoke (control). Hence, with these obstacles posed by RCT, many researchers have resorted to the discovery and inference of causal structures from purely observational datasets, or a combination of both data and RCT [2], [3].

However, despite the successes of causal models using observational datasets, many of their designs remain untested or unvalidated for correctness, according to the extant literature. A recent study by Tennant and Murray [4], which investigated model testing in the healthcare sector, revealed that none of the 200 reviewed articles were tested or validated for correctness. Thus, if these models are to be further used in the estimation or evaluation of causal inference of such projects, the estimation results may leave considerable room for dispute and doubt. Thus, in this study, we have designed an application-based novel SCM from the background knowledge gotten from the American university of Nigeria (AUN), Yola's project on the letter identification subtask of Early Grade reading Assessment (EGRA) program on Strengthen Education in Northeast Nigeria (SENSE), which was sponsored by the United State Agency for International Development (USAID), which occurred between 2021 to 2202. We employed the conditional independence test (CIT) criteria to test the correctness of our SENSE-EGRA SCM, and the results show a near-perfect model (See Table 1).

The main contribution of this work is as follows:

- Theoretical insight into the structural causal model (SCM) framework,
- Development of an application-based novel SCM for the SENSE-EGRA dataset
- Model correctness validation using the conditional independent test (CIT) criteria.
- Experiment Reproducibility. See the appendix links to data and CIT codes for reproduction of the experiment.

In Section 2, the basic theoretical concept of the causal model is discussed. Section 3 discusses direct acyclic graphs and their relations to causality and the Bayesian network factorization. Section 4 presents some of the main assumptions driving SCMs. Section 5 presents our experiment setup as it relates to the design of our SENSE-EGRA SCM. Section 6 presents our model correctness validation testing results using the CIT criteria. And finally, section 7 wraps up the study and gives direction on future work.

Fig. 1 SCM with (b) an intervention and without (a) an intervention

II. Materials and Methods

In this section, the various forms of causality are defined, followed by the two major frameworks used for causality, which are the structural causal model (SCM) framework and the potential outcome or Rubin causal model (RCM) framework, with a juxtaposition of both frameworks. The section concludes by explaining how causal interventions are executed in a dataset using the SCM framework.

A. Causal Model

It is an abstraction of mathematics that describes quantitatively the relations of causality that exist among variables in an observable dataset [5]. These mathematical models are derived from the domain and background knowledge embodied in the DAG, and they evince the causal relations within the observable dataset [6]-[8].

B. Types of Causal Model

Two types of causal models exist for causality, which are (i) the Structural Causal Model (SCM) proposed by Pearl [7] and (ii) the Potential Outcome Framework, also called the Rubin Causal Model (RCM) [9], [10]. However, the study scope is limited to SCM and not the RCM. *An SCM:* The framework for causality based on SCM gives a holistic understanding of the theory of cause and effect. It is composed of two parts: the causal diagram (or graph) that encodes background domain knowledge and assumptions of the distribution (the dataset), and the Bayesian network factorization (BNF) or structural equations part, which models or algorithmized (mathematically) the relations among the study variables based on the causal assumptions from the graph [5], [11]-[13]. This work focuses more on SCM with a more detailed explanation of the connections of the graphs and the dataset in subsequent sections.

C. Causal Relations with SCM

Determining the causal relations that exist among variables in an observational study in a purely probabilistic distribution is an ambiguous and daunting task. If a conditional probability distribution such as $P(Y|X)$ for instance, represent the conditional probability distribution of obesity (Y) given a particular level of sugar intake (X). This distribution relation is ambiguous in terms of an experimental setting (RCT) where sugar intake was ascertained by randomization or merely through an observational process. In his book on causality, Pearl [7], to differentiate the mere conditional observational probability distribution (i.e., statistical association/correlation) and interventional conditional probability distribution (which is a causal association), introduced the do-operator of the do-calculus to differentiate interventional distribution from observational. Hence, the expression $P(Y|X)$ can now be regarded as a mere conditional observational association that depicts how the probability of Y (obesity) will change if someone were to observe the sugar intake (X). While $P(Y|do(X = x))$ is regarded as the interventional conditional probability distribution (which is a causal association), depicting the probability of obesity (Y) given that a measure unit of sugar (x) were taken (purposefully and not observed). Hence, making the observation and intervention distinct: $P(Y|X = x) \neq P(Y|do(X = x))$. The practical difference between the two may be the existence of a variable(s) Z (individual gene tar for instance) that may be confounding the relations, which exists in some back-door path: See figure 1 DAG for confounding relations. In the intervention distribution, the

causal effects is determined given difference values of the treatment/control X (i.e., when sugar is taken and when sugar is not taken) and this can be measured and compared in the interventional distribution, written as: $P(Y|do(X = 1))$, and $P(Y|do(X = 0))$ where 1 and 0 stands for treatment and no treatment (control) respectively for an individual instance, which is called the individual treatment effect (ITE). Thus, when this process involves all sampled or all cases of the population, the causal intervention is defined in terms of the average treatment effects (ATE) for the instances of the population. Written in terms of the expectation as: $\tau(1,0) = \mathrm{E}[\mathrm{Y} \mid \mathrm{do}(\mathrm{x} = 1)] - \mathrm{E}[\mathrm{Y} \mid \mathrm{do}(\mathrm{x} = 0)]$. Also, conditional treatment effects (CATE) can be taken for a subpopulation group in a similar manner as well. Thus, it can be seen that this kind of intervention model is the RCT experiment that determines causality in observational datasets [14], [15]. Despite Pearl's clear distinction between observational and interventional datasets [7], not all datasets can be neatly categorized into these two types, as some experiments may not clearly or wholly demonstrate the value of the variable that is intervened upon in the dataset. Thus, due to these two distinctions, which are obfuscated in the distributions, it has become imperative to represent causal models explicitly in terms of a directed acyclic graph (DAG) or simply a causal graph as proposed by Pearl [14]. Causal graphs in SCM are a crucial component that facilitates the identification of causality from the dataset; hence, we discuss them in the next section.

D. Causal Graph

This section presents causal graphs as applicable in SCM. Fundamental concepts in graph theory, such as the popular backdoor adjustment criteria and the Bayesian network factorization (BNF), are elicited and explicated.

E. Causal Graph Composition

A causal graph (denoted as $G = (V, E)$, consists of two or more nodes (also called vertices) representing a random variable set $(V), where\ V = X_1, X_2, X_3, \ldots X_n$ and a number of connecting lines among the nodes called edges (E). These random variables may include the observed and unobserved (if they exist) variables alongside the treatment and outcome variables. In Figure 2.1A, the graph is undirected because it lacks directional arrows on its edges. While 1B, the graph is directed because of the arrow direction. And 1C shows a directed graph with a cycle [16] and finally, 1D shows an intervention graph on the variable C. A directed edge from A to B (written as: A $\rightarrow B$) is interpreted as, B is caused by A or (A is the potential cause of B) [5]. Hence, with a causal graph, a hypothesized causal query can clearly be modelled through the causal pathways in the graph, and all dependent/independent relations as they relate to all variables associated with the query are known. And this graph model can be factorized using the Bayesian network factorization or the structural equations, based on some assumptions, to obtain a causal estimate and of the conditional probability distribution from which it can be used with the observed dataset to ascertain the causal estimate of the hypothesized query [14], [17].

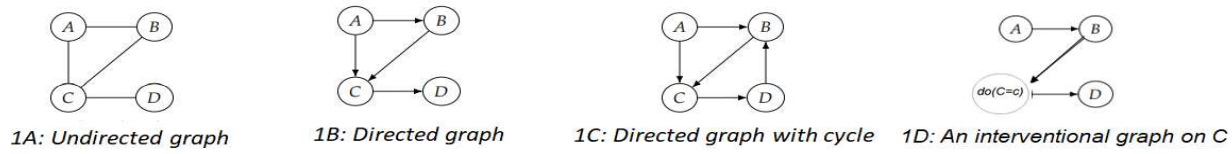


Fig. 2 Undirected, directed, directed with cycle, and intervention graph

A path in the graph is an oriented order of adjacent edges, irrespective of the direction of the adjoining nodes. For instance, $A - C - B$ are considered as a path in Figure 2.1A and $A \longrightarrow C \longleftarrow B$. There is also a path in Figure 2.1B. A directed path is one in which all edges are directed or pointing in the same direction, e.g., the path, $A \longrightarrow C \rightarrow B$. In Figure 2.1B, it is regarded as directed. Most causal algorithms work best with the directed acyclic graphs (DAGs) condition, as shown in Figure 2.1B, and a few causal algorithms work with the cyclic graph condition, as shown in Figure 2.1C [5], [11]-[13].

F. Three Cardinal Relations in Graphs

A descendant of a node A is a node $C \in V$, such that there is a direct edge from A to C (written as: $A \longrightarrow C$) in the DAG G. This corresponds to A being an ancestor (parent of) C. The progenies (A and B) of a node C, are the nodes in V with directed edges connecting C, (designated as: $A \longrightarrow C \longleftarrow B$). This child and the two parents' relationship is designated as $A \longrightarrow C \longleftarrow B$, is also called a collider [18], [19] or immorality [8], [16] is the first fundamental relation that can exist among variables represented in DAG. A second relation exists called a mediator or chain, where a parent node A (usually exogenous) that produces a child node C, where C in turn produces another child B (which is a grand descendant of A) [8], [14], [17]. finally, a third relationship exists where a node C, which is a single parent having two descendants A and B (written as: $A \longleftarrow C \longrightarrow B$) is called a fork or common cause confounder. Thus, these three relations (collider, chain/mediator, and fork) are the three common relations that exist in an observational dataset and can be mirrored or expressed in a DAG, forming the building block or structure in a causal graph for determining a relationship (causal or associational) in an observational setting [8], [11], [14], [17], [20], [21].

G. Causal Connection & the Backdoor Adjustment Criteria in a Graph

D-separation and d-connection are the processes that define a set of variables V's connectivity in a causal graph G [21]. The D in the d-separation and d-connection stands for dependency, and it is a process of establishing independence or dependency from two or more variables that are independent or otherwise on a third variable, cap C, in a DAG, which is a reflection in the dataset. For instance, in the case of a fork ($A \longleftarrow C \longrightarrow B$), or a chain/mediator ($A \longrightarrow C \rightarrow B$), the variable C is a link between both A and B. Hence, once you condition on the linking variable C, you will block or close the dependency relationship that exists between paths A and B. That is to say, paths A and B will become independent

conditioned on C, written as: $A \perp\!\!\!\perp B|C$. Albeit the reverse is the case, when it comes to the collider or immorality structure $(A \longrightarrow C \longleftarrow B)$, as the paths A and B are already independent or blocked in their current state (i.e., $A \perp\!\!\!\perp B \nmid C$: A is independent of B not conditioned on C), without the need for conditioning on any variable including C.

Hence, once you condition on C, a relationship between A and B is induced (i.e., A and B becomes dependent conditioned on C. written as: $A \not\perp\!\!\!\perp B|C$). This process of blocking the flow of unwanted association on non-causal pathways to determine causality only through a causal pathway is called the backdoor adjustment criterion [22], [23]. Pearl [21], defined the process of d-separation and d-connection for backdoor adjustment criteria in a DAG G formally as follows: A path connecting two variables A and B is said to be d-separated or blocked if and only if: (i) the path contains a fork such as : $(A \longleftarrow C \longrightarrow B)$ or chain/mediator such as: $(A \longrightarrow C \rightarrow B)$ that has been conditioned on C. Written as: (A$\perp\!\!\!\perp_G$ B|C), and (ii) the path between A and B contain a collider on C, such as $(A \longrightarrow C \longleftarrow B)$ that has not been conditioned on, alongside any descendant of the collider C, that is not conditioned on as well. Written as: $(A \perp\!\!\!\perp_G B \nmid C)$ or just $A \perp\!\!\!\perp_G B$. This same process of d-separation and the backdoor adjustment criteria from the graph G can be utilized to determine dependencies/independencies of variables in the distribution (or dataset), which is a factorization of the d-separation in the graph using the Bayesian Network Factorization (BNF). The d-separation in the distribution is written as: A$\perp\!\!\!\perp_p$ B|C, or A$\not\perp\!\!\!\perp_p$ B|C for independence and dependency conditions, respectively, similar to the d-separation in the graph with the subscript P to distinguish it from the graph's d-separation criteria, which is represented by the subscript G. This can further be used to determine causal relations in the distribution as a whole. On the other hand, a path from A and B through C, is said to be d-connected, unblocked, or open when it is not d-separated [17], [21].

H. The Bayesian Network Factorization (BNF) in Graphs

The DAGs are interpreted in two parts. i.e., the probabilistic and the causal interpretations. The probabilistic inference sees the directional arrows on the DAG G as showing probabilistic dependencies or associations among the variables of study, while the lack of arrows corresponds to the conditional independence asserted by the study variables [24]. Based on some assumptions, the simplest being the Markovian condition, which states that each study variable is considered independent of all its non-descendants in the graph except its direct parent. Usually written as $A \perp\!\!\!\perp B|C$. Hence, based on the assumption, the joint probability distribution function $P(v) = P(v_i, \dots, v_n)$ factorizes based on the BNF as:

$$P(v) = \prod_{i}^{n} P(v_i|pa_i) \tag{1}$$

where $v_i = 1, \dots, n, and \;\; pa_i$ denotes the parent of the variable $v_i$in the graph

Thus, based on the BNF of equation (1), the graph in Figure 2:1B, for instance, the probability distribution of it (i.e.,1B) can be factorized and summarized, based on the Markov assumption, as follows:

$$P(A, B, C) = P(A)P(B|A)P(C|B, A)P(D|C) \tag{2}$$

This contrasts with the normal Bayesian probability distribution network, which uses the chain rule without the graph and the Markov assumption, written as:

$$P(A, B, C) = P(A)P(B|A)P(C|B, A)P(D|C, B, A) \tag{3}$$

The difference between equations (2) and (3) is in the last product, the conditional probability of D, where equation (2) reduces the conditioning probability to only its immediate parent node C, based on the position of equation (1) and as captured in the graph of Figure 2:1B. While equation (3) assumes no graph and factorizes the distribution using the chain rule. Hence, the probability of D, given (or conditioned on:) C, B and A are used as elicited in equation (3).

I. Causal Identifiability with BNF Intervention Graphs

The second interpretation of the graph is called a causal interpretation. In this scenario, the arrow's direction in the DAG G represents the influence of causality among the variables. Here, the BNF of equation (1) above is still essential, but the arrows are assumed to evince a separate process in the data generated. Hence, after eliciting a causal path from the DAG G, the conditional probability of the distribution $P(v_i|pa_i)$, which is generated based on the graph G, and which is a statistical estimand, can be estimated from the data. The relations of conditional dependency expressed by the BNF formula of equation (1) do not necessarily lead to causal inference (due to the mixtures of confounding variables sometimes).

However, equation (1) can be extended to cater for interventions (which are causal in their implementation) as presented by [7]. Using the do-operator of the do-calculus as an intervention on the desired variable (or node) the difference between mere conditional distribution (correction), written as: $P(Y|X = x)$, and the causal intervention of the conditional distribution, written as: $P(Y|do(X = x))$, in the graph and subsequently the data can be clearly distinguished. For instance, if the graph in Figure 2 were derived from the query hypothesis of determining the effects of shoe size X on the reading ability Y of children. The age variable Z, confounds the relationship between reading ability Y and shoe size X, making them have a statistical correlation as shown in Figure 1(a). But when you carry out an intervention on the shoe size X such as $P(Y|do(X = x))$, the age variable Z that confounds the relations is severed, and the conditional probability of the BNF produces an estimand which is given as $P\big(Y\big|do(X = x)\big) = P(Z)P(X|Z)P(Y|Z, X)$. Which is summarized by getting rid of the factor for probability of X in the BNF to get: $P\big(Y\big|do(X = x)\big) = \sum_z P(Y|Z, X)\, P(Z)$. With this causal intervention estimand, using the d-separation and the backdoor criteria, the shoe size X will be set to a treatment unit of 1 and no treatment (control) unit of 0, while conditioning on a certain age Z say 8years.

Thus, the difference between the treatment and no treatment of the shoe size $(X: 0,1)$ generated from conditioning on a certain age $(Z = 8)$ for the set of $\boldsymbol{Z}$ variable in the dataset can be calculated as the ATE, given mathematically in terms of their expectation as: $\tau(1,0) = E[Y|do(x = 1)] - E[Y|do(x = 0)]$, which translates to the

causal estimand or causal inference estimation on the effect of shoe size X on reading ability Y in children. This estimand would likely be zero (no effect), thus killing the lurking variable (age) and exposing the spurious association (correlation) that exists between shoe size X and reading ability Y. Note, however, that if the confounding variable Z is unobserved or not part of the distribution (the dataset), the causal identification of X on Y It cannot be feasible to obtain from the data, even though it is revealed in the graph. This do-operator, which translates to intervention and causality in data, differentiates from mere association (correlation) that is used in machine learning algorithms.

With SCM, counterfactual hypothesized queries, which are carried out on an individual level of the sampled dataset, can also be estimated, using some techniques proposed by [25], [26], which transcend the do-operator of the do-calculus, which only work with the i.i.d. condition [27]. Although counterfactual causal effects would not be covered in this work.

J. Assumptions in SCM

This section covers the three major assumptions often used for causality, especially with i.i.d. datasets, thus driving the process of causality in observational data settings with the SCM framework. These assumptions are: (1) The Markov assumption, (2) The Acyclicity assumption, and (3) The causal sufficiency assumption. These assumptions are summarized as follows:

K. The Markov Assumption

This assumption states that a parent node in a DAG G Representing a variable is considered independent of all its non-descendants in the graph, except its direct parent. This assumption ensures that the causal estimand for identifying causal relations is generated from the graph to the data, using the BNF or the structural equation of the functional causal model (FCM). This estimand, which is modeled using the Markov condition when it is sufficient (i.e., all confounding variables identified), becomes the basis for estimating the probability distribution, a statistical estimand, from the dataset. Equation (1) is a representation of the Markov condition. The Markov assumption, when combined with the causal edge assumption, states that: in a DAG GAll adjacent nodes are dependent; this can generally be referred to as the minimality assumption [10], [16], [28].

L. The Acyclicity Assumption

It is the phenomenon that ensures that the set of adjoining variable nodes $\boldsymbol{V}$ in the causal graph, it does not form a cycle, a feedback loop, or go back in time, as shown in Figure 2:1C, but are rather directed and acyclic, as shown in Figure 2.1B B [29], [30].

M. The Causal Sufficient Assumption

This condition states that in a given causal graph GThere are no variables confounding relationships that are unobserved among the study variables. That is to say, the causal sufficiency assumption ensures that all variables that may be confounding or having a hidden effect on the hypothesized query variable of treatment and outcome (t, y) are identified and explicitly shown on the graph, whether or not they are observed in the distribution of the dataset [31]-[33]. Hence, these are the assumptions that are employed in the development of our SENSE-EGRA SCM.

III. Results and Discussion

The EGRA (SENSE) dataset, focusing on the subtask of letter identification for grade 2 students in two northeast states of Nigeria, comprises 1,114 records collected from over 200 primary schools in these states. Nineteen columns are of interest for our design of the SCM and analysis. These columns are further grouped into five distinctive groups, which are: A set of input features or covariates (X) where X stands for $State, LGA, Gender, Age$ etc., the output feature LI_3 (Y), the treatment variable T $(Treament)$ and two other assessment or evaluation features $(LI\_1,$ and $LI\_2)$ respectively. See the appendix for more details on the dataset-encoded meanings.

Thus, based on the above-discussed methodology in section 2, we designed the EGRA- SENSE SCM of Figure 3, and validated for correctness with the dataset as shown in Equation 4 below, and the result is presented in Table 1:

$$\begin{aligned} &LI\_1 \perp X | LI\_2, T \\ &LI\_2 \perp T | X \\ &LI\_3(Y) \perp T | LI\_1, LI_2, \\ &LI\_3(Y) \perp X | LI\_2, T \\ &LI\_3(Y) \perp X | LI\_1, LI\_2, \end{aligned} \tag{4}$$

Thus, the estimand and the back-door adjustment criteria, which identified the admissible set of covariates required for adjustment in our EGRA- SENSE SCM, as shown in Figure 3(a), are given as:

$$\mathrm{P(T, X, LI\_2, LI\_3) = P(LI\_3 | X, LI\_2, T)} \tag{5}$$

And the corresponding NPSEM generated from the mutilated DAG, as shown in Figure 3(b), for our SENS-EGRA SCM designating an intervention distribution is given as:

$$\begin{aligned} x = \mathrm{f}_x(U_x), \quad t = t', \\ li_2 = \mathrm{f}_{li_2}(x, U_{li_2}), li_1 \\ = \mathrm{f}_{li_2}(t, U_{li_1}), li\_3 \\ = \mathrm{f}_{li_3}(li_1, li\_2, U_{li\_3}) \end{aligned} \tag{6}$$

Notice that LI_1 is not conditioned on, since from the DAG, it is considered a post-treatment or mediator variable. Some studies by [7], [24], [25], [34], advised against conditioning on such post-treatment or mediator variables. Section 6, presents the result of the conditional independence test (CIT) implemented in an R package called Daggity [35].

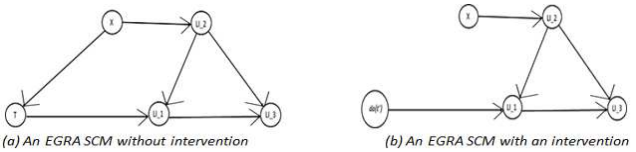
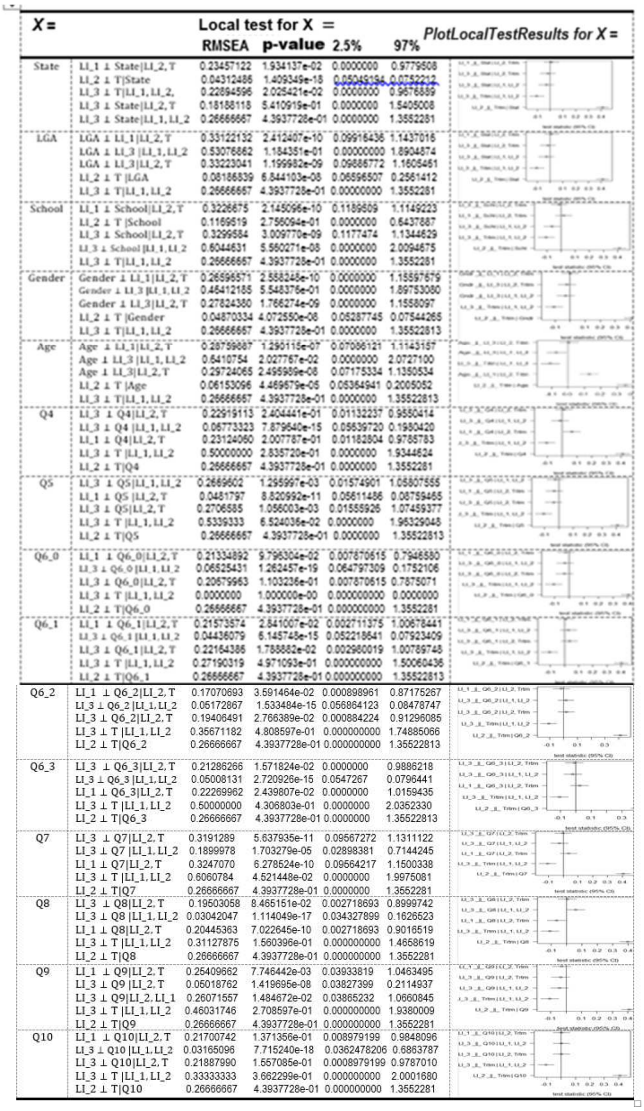


Fig. 3 EGRA- SENSE- SCM with (b) and without (a) intervention

A. CIT Results for SENSE-EGRA SCM

SCM is a qualitative process that is subjective based on background knowledge. Hence, experts advise validation and testing of the model with the dataset to ensure its correctness [4], [24], [25], [36]-[39]. One of the most pervasive validation

tests for SCM is the use of conditional independence testing (CIT) criteria [24], [25], [36]-[39]. Thus, once the validation process is over, the adjustment criteria can be applied to the SCM. Some studies by [24], [25], [37], proposed two adjustment criteria (the backdoor and front-door) depending on the structure of the SCMs in a concept called the d-separation (dependency separation). This concept, when properly applied to the SCM, is sufficient to identify the estimand (mathematical formula) for adjusting covariates and estimating the causal impact of the intervention. For our experiment, we implemented the CIT using the identified conditional independence set of equation 4 and applied the back-door adjustment criteria for eliminating confounding bias as shown in equations 5 and 6, respectively. Figure 4 below shows the result of the CIT performed on the dataset to verify and validate the correctness of our EGRA- SENSE SCM, implemented in the R package tool of [35].



X =		Local test for X = RMSEA	p-value	2.5%	97%	PlotLocalTestResults for X =
State	LI_1 ⊥ State|LI_2, T	0.23457122	1.934137e-02	0.0000000	0.9779508	
	LI_2 ⊥ T|State	0.04312486	1.409349e-18	0.05045196	0.0752212	
	LI_3 ⊥ T|LI_1, LI_2,	0.22894596	2.025421e-02	0.0000000	0.9676889	
	LI_3 ⊥ State|LI_2, T	0.18188118	5.410919e-01	0.0000000	1.5405008	
	LI_3 ⊥ State|LI_1, LI_2	0.26666667	4.3937728e-01	0.0000000	1.3552281	
LGA	LGA ⊥ LI_1|LI_2, T	0.35122132	2.412407e-10	0.09916436	1.1437016	
	LGA ⊥ LI_3 |LI_1, LI_2	0.53076862	1.184361e-01	0.00000000	1.8904874	
	LGA ⊥ LI_3|LI_2, T	0.33223041	1.199682e-09	0.09686772	1.1605461	
	LI_2 ⊥ T |LGA	0.08186939	6.844103e-08	0.06596507	0.2561412	
	LI_3 ⊥ T|LI_1, LI_2	0.26666667	4.3937728e-01	0.00000000	1.3552281	
School	LI_1 ⊥ School|LI_2, T	0.3226675	2.145096e-10	0.1189809	1.1149223	
	LI_2 ⊥ T |School	0.1169519	2.756094e-01	0.0000000	0.6437687	
	LI_3 ⊥ School|LI_2, T	0.3299584	3.009770e-09	0.1177474	1.1344629	
	LI_3 ⊥ School |LI_1, LI_2	0.6044631	5.560271e-08	0.0000000	2.0094675	
	LI_3 ⊥ T|LI_1, LI_2	0.26666667	4.3937728e-01	0.0000000	1.3552281	
Gender	Gender ⊥ LI_1|LI_2, T	0.26996571	2.588245e-10	0.0000000	1.1859757	
	Gender ⊥ LI_3 |LI_1, LI_2	0.46412185	5.548376e-01	0.0000000	1.8975308	
	Gender ⊥ LI_3|LI_2, T	0.27824380	1.766274e-09	0.0000000	1.1558097	
	LI_2 ⊥ T |Gender	0.04870334	4.072550e-06	0.05287745	0.0754426	
	LI_3 ⊥ T|LI_1, LI_2	0.26666667	4.3937728e-01	0.0000000	1.3552281	
Age	Age ⊥ LI_1|LI_2, T	0.28759687	1.250115e-07	0.07096121	1.1143157	
	Age ⊥ LI_3 |LI_1, LI_2	0.6410754	2.027767e-02	0.0000000	2.0727100	
	Age ⊥ LI_3|LI_2, T	0.29724065	2.495989e-08	0.07175334	1.1350534	
	LI_2 ⊥ T |Age	0.06153096	4.469679e-05	0.05364941	0.2005052	
	LI_3 ⊥ T|LI_1, LI_2	0.26666667	4.3937728e-01	0.0000000	1.3552281	
Q4	LI_3 ⊥ Q4|LI_2, T	0.22919113	2.404441e-01	0.01132237	0.9580414	
	LI_3 ⊥ Q4 |LI_1, LI_2	0.06773323	7.879540e-15	0.05639720	0.1980420	
	LI_1 ⊥ Q4|LI_2, T	0.23124060	2.007787e-01	0.01182804	0.9785783	
	LI_3 ⊥ T |LI_1, LI_2	0.50000000	2.835720e-01	0.0000000	1.9344624	
	LI_2 ⊥ T|Q4	0.26666667	4.3937728e-01	0.0000000	1.3552281	
Q5	LI_3 ⊥ Q5|LI_1, LI_2	0.2669602	1.256997e-03	0.01574901	1.05807555	
	LI_1 ⊥ Q5 |LI_2, T	0.0481797	8.820992e-11	0.05611486	0.08759455	
	LI_3 ⊥ Q5|LI_2, T	0.2706585	1.056003e-03	0.01558926	1.07459377	
	LI_3 ⊥ T |LI_1, LI_2	0.5339333	6.524036e-02	0.0000000	1.96329048	
	LI_2 ⊥ T|Q5	0.26666667	4.3937728e-01	0.0000000	1.3552281	
Q6_0	LI_1 ⊥ Q6_0|LI_2, T	0.21334892	9.796304e-02	0.007870615	0.7948580	
	LI_3 ⊥ Q6_0 |LI_1, LI_2	0.06525431	1.262457e-19	0.064797309	0.1752106	
	LI_3 ⊥ Q6_0|LI_2, T	0.20679963	1.103236e-01	0.007870615	0.7875071	
	LI_3 ⊥ T |LI_1, LI_2	0.0000000	1.000000e-00	0.0000000	0.0000000	
	LI_2 ⊥ T|Q6_0	0.26666667	4.3937728e-01	0.0000000	1.3552281	
Q6_1	LI_1 ⊥ Q6_1|LI_2, T	0.21573574	2.841007e-02	0.002711375	1.00678441	
	LI_3 ⊥ Q6_1 |LI_1, LI_2	0.04436079	5.145745e-15	0.052218641	0.07923409	
	LI_3 ⊥ Q6_1|LI_2, T	0.22164386	1.788682e-02	0.002980019	1.00789748	
	LI_3 ⊥ T |LI_1, LI_2	0.27190319	4.971093e-01	0.0000000	1.50060436	
	LI_2 ⊥ T|Q6_1	0.26666667	4.3937728e-01	0.0000000	1.3552813	
Q6_2	LI_3 ⊥ Q6_2|LI_2, T	0.17070693	3.591464e-02	0.000898961	0.87175267	
	LI_3 ⊥ Q6_2 |LI_1, LI_2	0.05172867	1.533484e-15	0.056864123	0.08478747	
	LI_3 ⊥ Q6_2|LI_2, T	0.19406491	2.766389e-02	0.000884224	0.91296085	
	LI_3 ⊥ T |LI_1, LI_2	0.35671182	4.808597e-01	0.0000000	1.74885066	
	LI_2 ⊥ T|Q6_2	0.26666667	4.3937728e-01	0.0000000	1.35522813	
Q6_3	LI_3 ⊥ Q6_3|LI_2, T	0.21286266	1.571824e-02	0.0000000	0.9886218	
	LI_3 ⊥ Q6_3 |LI_1, LI_2	0.05008131	2.720926e-15	0.0547267	0.0796441	
	LI_1 ⊥ Q6_3|LI_2, T	0.22269962	2.439807e-02	0.0000000	1.0159435	
	LI_3 ⊥ T |LI_1, LI_2	0.50000000	4.306803e-01	0.0000000	2.0352330	
	LI_2 ⊥ T|Q6_3	0.26666667	4.3937728e-01	0.0000000	1.35522813	
Q7	LI_3 ⊥ Q7|LI_2, T	0.3191289	5.637935e-11	0.09567272	1.1311122	
	LI_3 ⊥ Q7 |LI_1, LI_2	0.1899978	1.703279e-05	0.02898381	0.7144245	
	LI_1 ⊥ Q7|LI_2, T	0.3247070	6.278524e-10	0.09564217	1.1500338	
	LI_3 ⊥ T |LI_1, LI_2	0.6060784	4.521448e-02	0.0000000	1.9975081	
	LI_2 ⊥ T|Q7	0.26666667	4.3937728e-01	0.0000000	1.3552281	
Q8	LI_3 ⊥ Q8|LI_2, T	0.19503058	8.465151e-02	0.002718693	0.8999742	
	LI_3 ⊥ Q8 |LI_1, LI_2	0.03042047	1.114049e-17	0.034327899	0.1626523	
	LI_1 ⊥ Q8|LI_2, T	0.20445363	7.022645e-10	0.002718693	0.9016519	
	LI_3 ⊥ T |LI_1, LI_2	0.31127875	1.560396e-01	0.0000000	1.4658619	
	LI_2 ⊥ T|Q8	0.26666667	4.3937728e-01	0.0000000	1.3552281	
Q9	LI_1 ⊥ Q9|LI_2, T	0.25409662	7.746442e-03	0.03933819	1.0463495	
	LI_3 ⊥ Q9 |LI_2, T	0.05018762	1.419895e-08	0.03827399	0.2114937	
	LI_3 ⊥ Q9|LI_2, LI_1	0.26071557	1.484672e-02	0.03865232	1.0660845	
	LI_3 ⊥ T |LI_1, LI_2	0.46031746	2.708597e-01	0.0000000	1.9380009	
	LI_2 ⊥ T|Q9	0.26666667	4.3937728e-01	0.0000000	1.3552281	
Q10	LI_1 ⊥ Q10|LI_2, T	0.21700742	1.371356e-01	0.008979199	0.9848096	
	LI_3 ⊥ Q10 |LI_1, LI_2	0.03165096	7.715240e-18	0.0362478206	0.6863787	
	LI_3 ⊥ Q10|LI_2, T	0.21887990	1.557085e-01	0.0008979199	0.9787010	
	LI_3 ⊥ T |LI_1, LI_2	0.33333333	3.662299e-01	0.0000000	2.0001680	
	LI_2 ⊥ T|Q10	0.26666667	4.3937728e-01	0.0000000	1.3552281	

Fig. 4 The Result of The CIT Identified in Equation 10 for Each of The Variables X

B. CIT Results Discussion

When testing for conditional independence between two or more variables, it is required that their conditional dependency be zero [35]. Hence, with the use of the R tool of Ankur [35], as used in this work, the root mean square error of approximation (RMSEA) and the p-value results that are close to zero (our p-value threshold is set at 0.05) validate the assumptions evinced by the SCM. While the values of the RMSEA and p-value that deviate significantly from zero or that are statistically significant reveal the model's inaccuracy or lack of conditional dependency among them.

Thus, the R tool produced by Ankur [35], package functions LocalTests() and the PlotLocalTestResults() are used for the analysis of the CIT. The LocalTests() tests the CIT for each of the feature variables X under the five conditional independence conditions identified in our EGRA-SENSE SCM of equation 4 for the variable $X = State, LGA, Gender, Age$ etc., at a confidence interval of 95% for all test cases as shown in column 2 of Table 1. The PlotLocalTestResults() function plots the results of the localTests() function as shown in column 3. All the results indicate negative p-values and zero-scale RMSEA values. Thus, validating the correctness of our EGRA- SENSE SCM, as no conditional dependency exceeds 0.4 in all test cases, as shown in PlotLocalTestResults() graphical output in column 3 of the table, meaning their dependence is nearly zero.

IV. Conclusion

In this study, we have designed a novel application-based SCM from the background knowledge gotten from the American university of Nigeria (AUN), Yola's project on the letter identification subtask of Early Grade reading Assessment program on Strengthen Education in Northeast Nigeria (SENSE-EGRA), which was sponsored by the United State Agency for International Development (USAID), which occurred between 2021 to 2202. We employed the conditional independence test (CIT) criteria for the testing and validation of the models 'correctness, and the results show a near-perfect model. The main contribution of this work is in the explication of the theoretical insight into the structural causal model (SCM) framework, the development and correctness validation testing of an application-based novel SCM for the SENSE-EGRA dataset.

For future works, we shall use the developed SENSE-EGRA SCM alongside some adjustment and matching estimation techniques, such as ordinary least square regression adjustment, propensity score by (weighting, stratification and matching) to deal with confounding and selection biases to estimate the causal inference of SENSE-EGRA intervention program of the American University of Nigeria, Yola, Adamawa State, Nigeria under the sponsorship of USAID.

Conflict of Interest Declaration

All authors have no financial or proprietary interest in any material discussed in this work.

References

[1] J. F. Box, *R. A. Fisher, the Life of a Scientist*. New York, NY, USA: Wiley, 1978.

[2] K. Benson and A. J. Hartz, "A comparison of observational studies and randomized, controlled trials," *New Engl. J. Med.*, vol. 342, no. 25, pp. 1878–1886, Jun. 2000, doi: 10.1056/nejm200006223422506.

[3] S. L. Silverman, "From randomized controlled trials to observational studies," *Amer. J. Med.*, vol. 122, no. 2, pp. 114–120, Feb. 2009, doi:10.1016/j.amjmed.2008.09.030.

[4] P. W. Tennant *et al.*, "Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations," *Int. J. Epidemiol.*, vol. 50, no. 2, pp. 620–632, Apr. 2021, doi: 10.1093/ije/dyaa213.

[5] R. Guo *et al.*, "A survey of learning causality with data: Problems and methods," *ACM Comput. Surv.*, vol. 53, no. 4, pp. 1–37, Jul. 2020, doi:10.1145/3397269.

[6] C. Hitchcock and M. Rédei, "Reichenbach's common cause principle," in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., Winter 2020.

[7] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[8] J. Peters, D. Janzing, and B. Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms*. Cambridge, MA, USA: MIT Press, 2017.

[9] J. Neyman, "On the application of probability theory to agricultural experiments. Essay on principles. Section 9," *Statist. Sci.*, vol. 5, no. 4, pp. 465–472, 1990.

[10] D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies," *J. Educ. Psychol.*, vol. 66, no. 5, pp. 688–701, Oct. 1974, doi: 10.1037/h0037350.

[11] P. Spirtes, C. Glymour, and R. Scheines, "Discovery algorithms for causally sufficient structures," in *Causation, Prediction, and Search*. New York, NY, USA: Springer, 1993, pp. 103–162.

[12] S. Greenland, J. Pearl, and J. M. Robins, "Causal diagrams for epidemiologic research," *Epidemiology*, vol. 10, no. 1, pp. 37–48, Jan. 1999, doi: 10.1097/00001648-199901000-00008.

[13] S. L. Lauritzen, "Causal inference from graphical models," in *Complex Stochastic Systems*, O. E. Barndorff-Nielsen, D. R. Cox, and C. Klüppelberg, Eds. Boca Raton, FL, USA: Chapman and Hall/CRC, 2001, pp. 63–107.

[14] F. Eberhardt, "Introduction to the foundations of causal discovery," *Int. J. Data Sci. Anal.*, vol. 3, no. 2, pp. 81–91, Apr. 2017, doi: 10.1007/s41060-016-0038-6.

[15] J. Y. Halpern, "Review of 'The Book of Why: The New Science of Cause and Effect', by Judea Pearl and Dana Mackenzie," *J. Causal Inference*, vol. 7, no. 1, 2019, Art. no. 20190013, doi: 10.1515/jci-2019-0013.

[16] B. Neal, "Introduction to causal inference from a machine learning perspective," 2020, *arXiv:2007.07041*.

[17] F. Elwert, "Graphical causal models," in *Handbook of Causal Analysis for Social Research*, S. L. Morgan, Ed. Dordrecht, Netherlands: Springer, 2013, pp. 245–273.

[18] A. R. Nogueira *et al.*, "Methods and tools for causal discovery and causal inference," *WIREs Data Mining Knowl. Discov.*, vol. 12, no. 2, p. e1449, Mar. 2022, doi: 10.1002/widm.1449.

[19] L. Yao *et al.*, "A survey on causal inference," *ACM Trans. Knowl. Discov. Data*, vol. 15, no. 5, pp. 1–46, May 2021, doi:10.1145/3444944.

[20] C. Glymour, K. Zhang, and P. Spirtes, "Review of causal discovery methods based on graphical models," *Front. Genet.*, vol. 10, p. 524, Jun. 2019, doi: 10.3389/fgene.2019.00524.

[21] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo, CA, USA: Morgan Kaufmann, 1988.

[22] L. Gultchin *et al.*, "Differentiable causal backdoor discovery," in *Proc. 23rd Int. Conf. Artif. Intell. Statist. (AISTATS)*, Palermo, Italy, 2020, pp. 1–11.

[23] J. Correa and E. Bareinboim, "Causal effect identification by adjustment under confounding and selection biases," in *Proc. 31st AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, 2017, pp. 3740–3746.

[24] J. Tian and J. Pearl, "A general identification condition for causal effects," in *Proc. 18th Nat. Conf. Artif. Intell. (AAAI)*, Edmonton, AB, Canada, 2002, pp. 567–573.

[25] J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*. New York, NY, USA: Basic Books, 2018.

[26] J. Pearl, "Theoretical impediments to machine learning with seven sparks from the causal revolution," in *Proc. 11th ACM Int. Conf. Web Search Data Mining (WSDM)*, Marina Del Rey, CA, USA, 2018, pp. 3–3.

[27] T. S. Richardson and J. M. Robins, "Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality," Center Statist. Social Sci., Univ. Washington, Seattle, WA, USA, Tech. Rep. 128, 2013.

[28] J. Zhang and P. Spirtes, "Intervention, determinism, and the causal minimality condition," *Synthese*, vol. 182, no. 3, pp. 335–347, Oct. 2011, doi: 10.1007/s11229-010-9751-1.

[29] A. Hauser and P. Bühlmann, "Characterization and greedy learning of interventional Markov equivalence classes of directed acyclic graphs," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 2409–2464, Aug. 2012.

[30] A. Hauser and P. Bühlmann, "Jointly interventional and observational data: estimation of interventional Markov equivalence classes of directed acyclic graphs," *J. Roy. Stat. Soc. B*, vol. 77, no. 1, pp. 291–318, Jan. 2015, doi: 10.1111/rssb.12071.

[31] R. Mayrhofer and M. R. Waldmann, "Sufficiency and necessity assumptions in causal structure induction," *Cogn. Sci.*, vol. 40, no. 8, pp. 2137–2150, Nov. 2016, doi: 10.1111/cogs.12318.

[32] J. Zhang and W. Mayer, "Weakening faithfulness: some heuristic causal discovery algorithms," *Int. J. Data Sci. Anal.*, vol. 3, no. 2, pp. 93–104, Apr. 2017, doi: 10.1007/s41060-016-0033-y.

[33] J. Zhang and P. L. Spirtes, "Strong faithfulness and uniform consistency in causal inference," *arXiv:1212.2506*, Dec. 2012.

[34] I. Shpitser, T. VanderWeele, and J. M. Robins, "On the validity of covariate adjustment for estimating causal effects," *arXiv:1203.3515*, Mar. 2012.

[35] A. Ankan, I. M. Wortel, and J. Textor, "Testing graphical causal models using the R package 'dagitty'," *Curr. Protoc.*, vol. 1, no. 2, p. e45, Feb. 2021, doi: 10.1002/cpz1.45.

[36] F. Thoemmes, Y. Rosseel, and J. Textor, "Local fit evaluation of structural equation models using graphical criteria," *Psychol. Methods*, vol. 23, no. 1, pp. 27–41, Mar. 2018, doi:10.1037/met0000147.

[37] J. Pearl and T. S. Verma, "A theory of inferred causation," in *Studies in Logic and the Foundations of Mathematics*, J. F. A. K. *et al.*, Ed. Amsterdam, Netherlands: Elsevier, 1995, vol. 134, pp. 789–811.

[38] J. Pearl and E. Bareinboim, "Transportability of causal and statistical relations: A formal approach," in *Proc. 25th AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, 2011, pp. 1–8.

[39] J. Pearl, "Causal inference in statistics: An overview," *Statist. Surv.*, vol. 3, pp. 96–146, 2009, doi: 10.1214/09-SS057.