JCTA Journal of Computing Theories and Applications ISSN:3024-9104

Research Article

Design and Validation of Structural Causal Model: A Focus on EGRA Dataset

Gabriel Terna Ayem [1,*], Ozcan Asilkan[1], and Aamo Iorliam[2]

[1] Computer Science Department, School of Information Technology and Computing, American University of Nigeria, Yola, Nigeria. e-mail : gabriel.ayem@aun.edu.ng, ozcan.asilkan@aun.edu.ng
[2] Data Science Department, School of Information Technology and Computing, American University of Nigeria, Yola, Nigeria; e-mail: aamo.iorliam@aun.edu.ng
* Corresponding Author: Gabriel Terna Ayem

Abstract: Designing and validating structural causal model (SCM) 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 (DAG), a component of an SCM, without testing them. This phenomenon is worrisome and is bound to cast a shadow on the inference estimates that may arise from such models. In this study, we have designed a novel application-based SCM for the first time using the background knowledge obtained from the Early Grade Reading Assessment (EGRA) program called the Strengthen Education in Northeast Nigeria (SENSE-EGRA), which is an educational intervention program of the American University of Nigeria (AUN), Yola, on the letter identification subtask. This project was sponsored by the United States Agency for International Development (USAID). We employed the conditional independence test (CIT) criteria for the validation of the SCM's correctness, and the results show a near-perfect SCM.

Keywords: Structural causal models; Casal validation; Conditional independent test; Observational datasets; EGRA.

1. 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 effect; and not having equipment, enough facts, or the sine-quo-non to ascertain this phenomenon of knowing what actions (causes) that produces the effects especially 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. This, drove the savaged man from his initial state of higgledy-piggledy to embrace the practice of magic, astrology, and certain fetish ways to achieve the causation phenomenon to overcome his bewildered state. Gradually, as societies evolved and advanced, mankind himself advanced from the primitive and savaged state to the current state of scientific and technological advancement of today's world. Thus establishing his hegemony on earth over and above every other species. Thus, 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 wanes to scientific logic, and astrology metamorphosed into astronomy and other technological innovations such as computer predictions, simulations, etc., became 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. The traditional scientific way of ascertaining this act is resident with the randomized controlled experiment or randomized controlled trial (RCT) method. This RCT method and idea is credited to Fisher [1]. Thus, this standard framework for causal discovery known as RCT always involves setting some (usually half) of the sampled

Received: October, 2nd 2023
Revised: November, 11th 2023
Accepted: November, 14th 2023
Published: November, 17th 2023

population of study and giving them a treatment (an intervention) under the same conditions, while the second half of the study population is left untreated (not intervened on) or controlled under the same or similar conditions. This approach helps 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 perform an RCT on a hypothesized query that seeks to uncover the health benefits or otherwise of smoking on a certain 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 others not to smoke (controlled). Hence, with this obstacle posed by RCT, many researchers have resorted to the discovery and inferring of causality from purely observational datasets (using SCM or the potential outcome (PO) framework), or a combination of both data and RCT [2, 3]. These frameworks (SCM & PO) have techniques that simulate RCT by dealing with the issues of confounding and selection biases in order to infer causality from observational datasets.

However, despite the successes recorded by causal models using observational datasets, many SCM designs are not tested or validated for correctness as far as the extant literature would reveal. A recent study by Tennant et al. [4], which investigated SCM DAG testing in the healthcare sector, evinced that among the 200 articles reviewed that relates to SCM design, not a single one of them was 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 a lot of room for dispute and doubt over the issue of effectively dealing with confounding bias. Thus, the purpose of having a well-designed and validated SCM includes [5-8]: (i) A well-designed and validated direct acyclic graph (DAG) formation for an SCM is a perfect representation of the data-generating process of an intervention program and this process is further employed in the estimation of the intervention effects for the intervention program (ii) an SCM with correct DAG formation will help in the identification of the right structural equations and covariates set to perform adjustment on to effectively handle the issue of confounding bias in the dataset during inference or impact estimation. Thus, a wrong SCM will have incorrect DAG formation and lead to the identification of the wrong set of structural equations and covariates set to perform adjustment on, and concomitantly bring about a biased estimation of the impact estimation for the intervention program, (iii) Finally, any inference estimates from an incorrect DAG formation for an SCM, representing an intervention dataset is not free of confounding and selection biases. Overall, it is worthy of note that all dataset analysis without an SCM DAG formation can only reveal association (or correlation), and these correlations are sometimes casual in nature while at other times, they are completely spurious. In the case where the relations between variables are spuriously correlated, it is the correct DAG formations of the variables involved that can identify which variable is a confounder (the one bringing about the spurious relations). And when the confounding variable is identified and removed or adjusted upon, the spurious relationship is severed. Take, for instance, scenario A, where ice cream sales have a positive correlation with high temperature. That is, the higher the temperature, the more sales that are recorded with ice cream. In scenario B, where electricity prices have a positive correlation with temperature. That is, the higher the temperature, the more prices one pays for electricity and the verse-verse. In scenario A, the correlation seems causal in nature, while in scenario B it is completely spurious. If one collects data on these three variables for each of the scenarios, i.e., ice-cream quantity supplied (let's call it X), ice-cream sales prices (let's call it Y), temperature of the day (let's call it Z) in the case of scenario A.

Similarly, in scenario B, we collect data for electricity consumption (called X), Prices per consumption (called Y), and temperature of the day (called Z). The DAG formation of Figure 1 aptly captures the relations of these two scenarios. In scenarios A and B, if one isolates the variable Z (perform an adjustment on it), one will realize that X and Y have little or no relationship. Thus, it is pertinent to know that data analysis without these DAG formations of an SCM depicting the relations of these variable sets cannot reveal the truths painted in both of these scenarios. The effective Isolation of confounding to determine the actual relations between two variables is what RCT tests are good at achieving. Thus, the SCM framework achieves a similar feat by simulating this RCT experiment with datasets. Hence, it is extremely

important that the DAG formation of the SCM be validated with the dataset from the intervention program and found to be correct in order to be assured of reliable inference estimates from such a model.

Other methods of testing and validating datasets exist for ML and other causal evaluation methodologies, such as Granger causality, cross-validation bootstrap validations, etc., [9, 10] but, they are not useful for validating the DAG structure of an SCM, which is our focused methodology in this study, and neither are they useful in terms of identifying the confounding variables, which causes confounding bias within the dataset. Take, for instance, the Granger causality, it is a causality test used popularly in the field of econometrics for testing causality with time series datasets [11, 12]. It uses a time series dataset that is similar or dissimilar in nature to predict and test another time series dataset. Cross-validation and bootstrapping tests are used in the validation of ML predictions in a dataset, and these sets of tests do not in any way reveal the dataset's data structure or the generating process of the dataset, which we seek to determine in this study. However, the CIT criteria of the SCM is capable of identifying the structure of an SCM DAG, which is a representation of the dataset structure and the confounding variables in the datasets, and it can sometimes be employed in dealing with selection bias as well [13, 14].

The aim of this study therefore is to show how a CIT validation process for an SCM can be achieved and performed with an intervention program dataset (SENSE-EGRA). Every intervention will require an evaluation of its impact, and this impact evaluation can only be correctly performed on the dataset that has being correctly validated for correctness using the CIT criteria for an impact evaluation with the SCM framework. Thus, in doing so, we used the intervention dataset that is obtained from the American University of Nigeria (AUN), Yola's project on the letter identification subtask for the Early Grade Reading Assessment (EGRA) program called Strengthen Education in Northeast Nigeria (SENSE-EGRA), which was sponsored by USAID. This program was conducted between 2021 to 2202. Thus, with this dataset, we designed an application-based novel SCM from the background knowledge of the dataset and thus, using the CIT criteria for the correctness validation of the designed SCM, the results show a near-perfect model. See Table 4 and Figures 3, 4, and 5 for the conceptual framework procedure, SENSE-SCM model design, and the CIT results.

1.1 Related Works

Other similar EGRA projects have used different techniques for the analysis and estimation of some EGRA intervention programs as it concerns the task of letter identification and other tasks in different parts of the world, as shown in these references [8, 15] [16], [17], [18], [19], [20-22]. See Table 1 for a detailed summary of the related works of those references.

Table 1. Shows summaries of the related works on EGRA Studies across the world on Letter Identification and Other Tasks with their Methods of Evaluations

Ref	EGRA Intervention Tasks/Grade Level	Evaluation Method	EGRA Name/ Country
[8, 15]	Letter identification and 6 other tasks/2	SCM and PO	SENSE/Nigeria
[23]	Letter Identification and 4 other tasks/1-9	PO and Bayesian Additive Tree (BART)	Arabic Assessment/ Lebanon and Syria
[16]	Letter identification and 6 others/1-3	Descriptive and Inferential statistics	Kakuma and Kalobeyei/ Kenya
[17]	Letter recognition/1-3	RCT & Descriptive/ Inferential statistics	Improving Reading/ South Africa
[18]	Letter identification & Mathematics /1 & 2.	Quasi-experiment & Descriptive/Inferential statistics	Learning for Living project: South Africa
[19]	Letter identification & Mathematics/1-9	Triple t-test Inferential statistics & Difference-in-Difference-in-Difference (DDD)	Literacy Program at the Right Age (Pacto pela Alfabetização na Idade Certa [PAIC]): Brazil
[20-22]	Letter Identification & 6 others/2 & 3	RCT & Descriptive and Inferential statistics	EGRA Plus: Liberia

However, a study by Oca et al. [23] implemented the potential outcome (PO) framework and the Bayesian Additive Regression Tree (BART) in an Arabic-EGRA intervention program task on letter identification and other subtasks without validating the framework. Thus, our study uses the SCM framework instead of the PO framework to design and validate SCM correctness in the area of letter identification for grade II students. Table 2 shows the major differences between our study and that of Oca et al. [23].

Table 2. Evince the differences between the study by Oca et al. [23] and ours as it concerns the methods, assumptions, and models employed

Comparison Indices	Oca et al. [23] Study	Our Study
Causal framework employed	PO Framework.	SCM Frameworks.
Assumptions used in overcoming confounding bias in covariates set	Unconfoundedness (ignorability), stable unit treatment value assumption (SUTVA), & and the Overlap.	SCM Backdoor adjustment criteria, with its do-calculus intervention process.
Model description of the dataset [Y:Yes/N: No]	N: No model description of the dataset is present.	Y: The model description of the dataset is coded in DAG. See Fig. 4.
Model & Assumptions' Validation [Y:Yes/N: No]	N: PO has no model, and its assumptions cannot be validated. Thus, there is a possibility of performing adjustment on the mediator variable, which can then cast a shadow on the causal impact estimates produced.	Y: Dataset assumptions encoded in the model (DAG) are validated using the CIT criteria. Thus validating the causal impact estimates that may come from the process. Fig. 3, 4, Table 4 & Algorithms 1 & 2 for model design, and model validation via CIT criteria.

1.2 Study Contributions, Study Structure, and Definition of Terms

This section discusses the contributions of this work and presents its structure and definitions of terms.

1.2.1 Study Contributions

The main contributions of this work are as follows:

1) Theoretical insight into structural causal model (SCM) framework,
2) Design of a conceptual framework for CIT criteria and impact evaluation processes.
3) Development of an application-based novel SCM for the SENSE-EGRA dataset
4) Designed of a general and specific algorithm procedure that can be used for the EGRA SCM models & the SENSE-EGRA model validation process.
5) Model validating using the conditional independent test (CIT) criteria
6) The empirical implementation of the experiment. See the Data Availability Statement link for the data and codes for the experiment reproducibility link.

1.2.2 Study Structure and Definition of Terms

In section 2, the basic theoretical concept of the structural 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. Finally, section 7 wraps up the study and gives direction on future work.

1.2.3 Study Definition of Terms

In this paper, just like is common with papers in the field, capital letters such as X represent a variable set. While their lowercase counterpart x, would represent instances of the variable set X. Also, characters such as T, Y, X_i would stand for single variables and their associates lower cases such as $t, y,$ and x_i would stand for their values respectively. Also, we use F_X or $f(X)$ for a function on a variable set X and an instance of such a function would be represented by F_x or $f(x)$. The calligraphic upper characters such $\mathcal{G}, \mathcal{V},$ and $\mathcal{E}$ stand for graph, node-set, and edges or vertices sets, respectively. For graphs of family relations, $Pa(V_i)$ stands for a set of parent nodes of a set of variables (V_i) found in the graph

and $pa(V_i)$ is an instance of $Pa(V_i)$. Similarly, the character $Ch(V_i)$, would stand for children node set in the graph $\mathcal{G}$ and the $ch(V_i)$ is an instance of $Ch(V_i)$. The letter T (its lowercase indicating its instance) is used as the treatment variable, and we assume the treatment to be binary and univariate. Similarly, the variable X is also used as a set of covariates in the graph. While the variable Y denotes the outcome variable with lower case, or lower case with subscript as an instance of it, or y with a bracketed binary digit such as $y_i(0), y_i(1)$ denotes instances of the treatment subscribed to them (which can also represent the potential outcome for treated and controlled under the PO framework). Finally, the symbol τ defines the various treatment effects, which is usually the change in the outcome variable for different treatment levels.

2 Basic Concept of Causal Models

In this section, the two major frameworks used for causality, which are the structural causal model (SCM) and the potential outcome or Rubin causal model (RCM) are defined; with particular emphasis on SCM techniques, DAGs formation and assumptions in the framework. Also, we present a brief juxtaposition explaining the major differences in both of these frameworks in Table 3, albeit major emphasis of this section is on the SCM framework.

2.1 Causal Model

A causal model is an abstraction of mathematics that describes quantitatively the relations of causality that exist among variables in an observable dataset [24]. These mathematical models are derived from the domain and background knowledge embodied in the DAG, and they evince the causal relations within the observable datasets [25-27].

2.2 Types of causal models

Two types of causal exist for causality, which are (i) the Structural causal model (SCM) proposed by Pearl [26] and (ii) the Potential outcome framework also called the Rubin causal model (RCM) [28, 29]. However, the study scope is limited to the SCM, and not the PO or RCM. Albeit Table 3 presents a brief juxtaposition of these two frameworks [30-32].

Table 3. A brief juxtaposition of these two frameworks for causal analysis in observational studies.

SCM Framework	PO Framework
Causal relations in the dataset are explicitly stated in the DAG and structural equations, which depict causality.	No depiction of causal relations in the dataset, rather, tables are used to represent potential outcomes of the subject under study with many missing data for counterfactuals.
The framework is model-driven and defines causality in terms of a single data generation Process (DGP).	The framework is data-driven and defines causality in terms of counterfactual and many DGP.
Variables that are not part of the dataset (e.g., instrument variable (IV)) but have causal relations with the dataset can also be represented in the DAG and factored in the inference estimation process.	Only variables in the datasets are factored in the inference estimation process since there is no DAG.
The framework uses theorems that are proven in the world to be true	The framework uses assumptions that have no proof in the real world
Confounding bias is dealt with using the backdoor adjustment criteria and, in rare cases, the front-door adjustment.	Confounding bias is dealt with using the unconfoundedness assumption.
The backdoor Adjustment criteria provide guidelines for how and where covariates adjustment can be made, e.g., no adjustment on mediators and colliders.	The unconfoundedness condition, which is the equivalent of the backdoor criteria, does not provide guidelines on how the adjustments are made. Adjustments are made based on the researcher's discretion.
Over-adjustment on covariates does not occur due to well-defined variable relations by DAG.	Over-adjustment on covariates can sometimes occur, which is capable of violating the overlap condition (selecting instances that are treated or controlled only and not both). This is a selection bias issue.

SCM Framework	PO Framework
The assumption encoded in the DAG, which enables the application of the backdoor adjustment criteria, can be validated in the dataset under the CIT criteria, as shown in the results of Table 4, and Figure 5 of our study.	The unconfoundedness assumption has no validation.
The do-calculus (do-operator) is used for intervention in SCM.	Intervention exists that is similar to the do-calculus but not explicitly stated as do-calculus.
SCM is used mostly in the field of Computing and related disciplines.	PO is used mostly in social science and econometric disciplines.
SCM is best suited when the goal is to learn the causal relations of variables in the dataset.	PO is best suited when the goal is to quickly estimate the effects of a given treatment on some outcome, which is the causal inference, and the emphasis is not on the causal relations.
SCM was proposed by Judea Pearl, a Computer Scientist	The PO framework was proposed by Donald Rubin

2.2.1 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, 24, 33, 34]. This work focuses more on the SCM with a more detailed explication of the connections between the graphs and the dataset in subsequent sections.

2.3 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, represents 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 [26] differentiated the mere conditional observational probability distribution (i.e., statistical association/correlation) from the interventional conditional probability distribution (which is a causal association), and introduced the do-operator or the do-calculus to differentiate interventional distribution from observational'. Hence, the expression $P(Y|X)$ can now be regarded as a mere conditional observational 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 measured unit of sugar (x) were taken (purposefully and not observed). Hence, making the observation and intervention distinct: $P(Y|X = x) \neq P\big(Y\big|do(X = x)\big)$.

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 are 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 $P\big(Y\big|do(x = 1)\big)$, and $P\big(Y\big|do(x = 0)\big)$ 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 instances 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 $\tau(1,0) = E[Y|do(x = 1)] - E[Y|do(x = 0)]$.

(a) (b)

Figure 1. Depicts the observational statistical correlational relations distribution (a) An SCM without intervention (b) An SCM under the intervention dot(t)

Also, conditional average treatment effects (CATE) can be similarly taken for sub-group populations as well. Thus, it can be seen that this kind of intervention models or simulates an RCT experiment that determines causality in the observational dataset [35, 36]. Despite the clear distinction describing and differentiating these two processes by Pearl et al. [26], not every dataset can be neatly categorized into this distinction between observational and interventional datasets, as some experiments may not clearly or wholly show the value of the variable that is intervened on 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 the directed acyclic graph (DAG) or simply causal graph as proposed by Pearl et al. [35]. The causal graph in SCM is an essential component that makes it easier to identify the causality from the dataset; hence, we discuss them in the next section.

3 Causal Graph

This section presents causal graphs as is applicable in SCM. Fundamental concepts in a graph such as the popular backdoor adjustment criteria and the BNF are elicited and explicated.

3.1 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 several connecting lines among the nodes called edges (E). These random variables may include the observed and existing (if they exist) variables alongside the treatment and outcome variables. Figure 2: 1A is an undirected graph due to the lack of directional arrows on them. While 1B, the graph is directed because of the arrow direction. And 1C shows a directed graph with a cycle [37], and finally, 1D shows an intervention graph on 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) [24]. Hence, with a causal graph, a hypothesized causal model can be designed through the causal pathways in the graph, and all dependent/independent relations as they relate to all variables associated with the query are known. This graph model can be factorized using the Bayesian network factorization or the structural equations; based on some assumptions to obtain a causal estimand of the conditional probability distribution from which it can be used with the observed dataset to ascertain the causal estimate of the hypothesized query [35, 38].

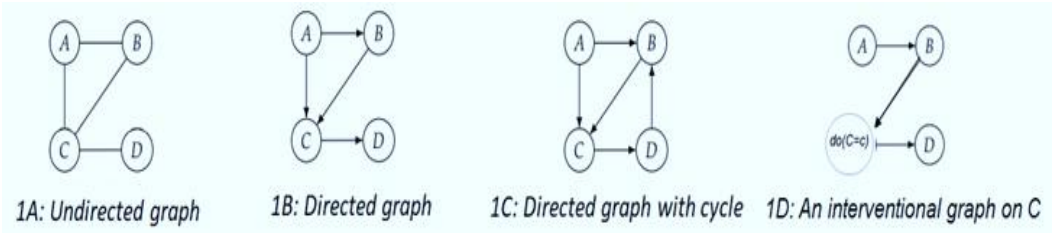


Figure 2. Shows an undirected, directed, directed with cycle, and intervention graph

A path in the graph is an oriented order of adjacent edges irrespective of the direction of the adjoining nodes. For instance, $A - C - B$ is considered as a path in Figure 2 1A and $A \longrightarrow C \longleftarrow B$ 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 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, 24, 33, 34].

3.2 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 a directed edge connecting C, (designated s: $A \longrightarrow C \longleftarrow B$). This child and two parents relationship designated as $A \longrightarrow C \longleftarrow B$ is also called a collider [39, 40] or immorality [27, 37] and is the first basic 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) produces a child node C, where in turn produces another child B (which is a grand descendant of A) [27, 35, 38]. 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 the causal graph for determining relationship (causal or associational) in observational settings [27, 33, 35, 38, 41, 42].

3.3 Causal Connection & the Backdoor Adjustment Criteria in a Graph

D-separation and d-connection are the processes that define a set of variable V's connectivity in a causal graph G [42]. 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 C in 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 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 $A \coprod B|C$. 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., $A \coprod B \nmid C$: 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 become dependent conditioned on C. written as $A \coprod B|C$). 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 criteria [43, 44]. Pearl et al. [42], 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: ($A \coprod_G B|C$), 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 collider C, that is not conditioned on as well. Written as: $(A \coprod_G B \nmid C)$ or just $A \coprod_G B$. 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 $A \coprod_p B|C$, or $A \coprod_p B|C$ 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 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 [38, 42].

3.4 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 the study, while the lack of arrows corresponds to the conditional independence asserted by the study variables [7]. 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 $A \coprod B|C$. Hence, based on the assumption, the joint probability distribution function $P(v) = P(v_i, \dots, v_n)$ factorizes based on the BNF as Equation (1).

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

Where $v_i = 1, \dots, n, and\ pa_i$ denotes the parent of the variable $v_i$in the graph [7, 24, 42].

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 Equation (2).

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

This contrasts the normal Bayesian probability distribution network, which uses the chain rule without the graph and the Markov assumption, written as Equation (3).

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

The difference in Equation (2) and (3) is in the last product 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).

3.5 Causal Identifiability with BNF Intervention Graphs

The second interpretation of the graph is called a causal interpretation. In this scenario, the arrow direction in the DAG G represents the causal relations 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, 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 to interventions (which are causal in their implementation) as presented by Pearl in [26]. 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 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 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 $P\big(Y\big|do(X = x)\big) = P(Z)P(X|Z)P(Y|Z, X)$. This is summarized by getting rid of the factor for the 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 8 years. Thus, the difference between the treatment and no

treatment of shoe size $(X: 0,1)$ generated from conditioning on a certain age $(Z = 8)$ for the set of 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 translate to the causal estimate or causal inference estimation on the effect of shoe size X on reading ability Y in children. This estimate 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 cannot be feasible to obtain in the data, even though it is revealed in the graph. This do-operator which translates to intervention and causality in data differentiates 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 Pearl [45, 46] which transcend the do-operator of the do-calculus, which only work with i.i.d condition [47]. Although counterfactual causal effects would not be covered in this work.

4. 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 setting with the SCM framework. These assumptions are (i) The Markov assumption, (ii) The Acyclicity assumption (iii) The causal sufficiency assumption. These assumptions are summarized as follows:

4.1 The Markov Assumption

This assumption states that a parent node in a DAG G representing a variable is considered independent of all its non-descendant in the graph except its direct parent. This assumption ensures that causal estimand for the identification of the causal relations is generated from the graph to the data, using the BNF or the structural equation of 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 which the probability distribution, which is a statistical estimand can be estimated from the dataset. Equation (1) is a representation of the Markov condition. The Markov assumption when combined with the causal edge assumption that states that: in a DAG G, all adjacent nodes are dependent; can generally be referred to as the minimality assumption [29, 37, 48].

4.2 The Acyclicity Assumption

It is the phenomenon that ensures that the set of adjoining variables nodes V in the causal graph 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 [49, 50].

4.3 The Causal Sufficient Assumption

This condition states that in a given causal graph G, there 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 have 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 [51-53]. Hence, these are the assumptions that are employed in the development of our SENSE-EGRA SCM.

5. Experiment Setup

This section explains the procedure for designing an SCM from the background and then identifying the criteria for performing a CIT with the designed SCM. It also explains the dataset and its focused task of letter identification, and how and when it was obtained and processed.

5.1 Dataset Description

According to a report from [54], two rounds of data collection were made (baseline & endline). The baseline assessment was carried out in November 2020 with 965 learners (482 from 146 schools in 11 local government areas (LGAs) in Adamawa, and 483 from 69 schools in 11 LGAs in Gombe). The end-of-project assessment was done in July 2021 with 964 learners (481 students from 125 schools in 11 LGAs in Adamawa, and 483 students from 70 schools in 11 LGAs in Gombe). Data is analyzed taking into account the sampling design (school strata, and sampling probability weights). The baseline sample contains a total of 457 boys and 508 girls, with an average age of 7.8 years. The end-of-project (endline) sample contains a total of 471 boys and 493 girls, with an average age of 8.8 years, making a total of 1,929 total datasets for the project.

5.2 Letter identification tasks

This task assesses a pupil's capability to identify the letter of an alphabet and its sounds naturally, without being hesitant. The task is made up of a page of a hundred upper/lowercase letters dispersed in ten rows of ten letters. These letters were randomly ordered. Also, the number of times each given letter appears is determined by the frequency that letter appears in primary school texts. The children were asked to say the letter and its sounds as many letters as they could in a minute. The evaluation score for this task is the number of letters a child correctly named in a minute. This measure is known as correct letter sounds per minute (CLSPM) identified in the dataset as LI_3 variable.

5.3 Initial Dataset Cleaning

The initial dataset cleaning performed according to a report from [54] was guided by the following checklist:

1) Review incomplete assessments.
2) Remove any "test" assessments that were completed before official data collection began.
3) Ensure that all assessments are linked with the appropriate school information for identification.
4) Ensure the child's assent was both given and recorded for each observation.
5) Ensure that all timed subtask scores fall within an acceptable and realistic range of scores.

5.4 Final Dataset Cleaning

After identifying the task to concentrate our analysis on (i.e., the letter identification subtask), we expunged the said subtask from the rest of the dataset in order to concentrate our analysis. Further, we then performed the second cleaning or data preprocessing on the dataset. Therefore, using the Python Jupiter notebook programing language, we were able to remove all the rows with missing values or NAs, and also removed the column with the student's ID/name from the dataset. Therefore, we ended up with a total of 1,114 records for the subtask of letter identification. For a detailed description of the dataset's variable names and their class options, please see the link in the Data Availability Statement. 19 columns are of interest for our design of the SCM and analysis. These columns are further grouped into 5 distinctive groups which are: A set of input features or covariates (X) where X instances are: $State, LGA, Gender, Age$ 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 link in the Data Availability Statement for more details on the dataset-encoded meanings.

Thus, based on the above-discussed methodology in section 2, we designed the SENSE-EGRA SCM of Figure 4 using the background knowledge generated from the data collection process, and validated the model's correctness with the dataset using the CIT criteria as shown in Equation (4), and the result is presented in Table 4. The entire CIT criteria and impact evaluation estimation process is shown in Figure 3. Further, in Figure 3, the dotted shapes show either an intangible process or an incomplete process in this study. Thus, the process of acquiring background knowledge of the EGRA task that generated the dataset is an intangible process and the processes of estimating the impact evaluation and interpretation of the impact evaluation are incomplete processes that are yet to be captured or completed in this study (See section 7.2 of our future work).

$$P(LI\_1 \perp X|LI\_2, T$$
$$LI\_2 \perp T|X$$
$$LI\_3(Y) \perp T|LI\_1, LI\_2 \quad (4)$$
$$LI\_3(Y) \perp X|LI\_2, T$$
$$LI\_3(Y) \perp X|LI\_1, LI\_2)$$

Where the term $\perp$ means independent of, and $|$ means given.

Thus, the estimand and the back-door adjustment criteria, which identified the admissible set of covariates required for adjustment in our SENSE-EGRA SCM, as shown in Figure 4(a) is given as Equation (5).

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

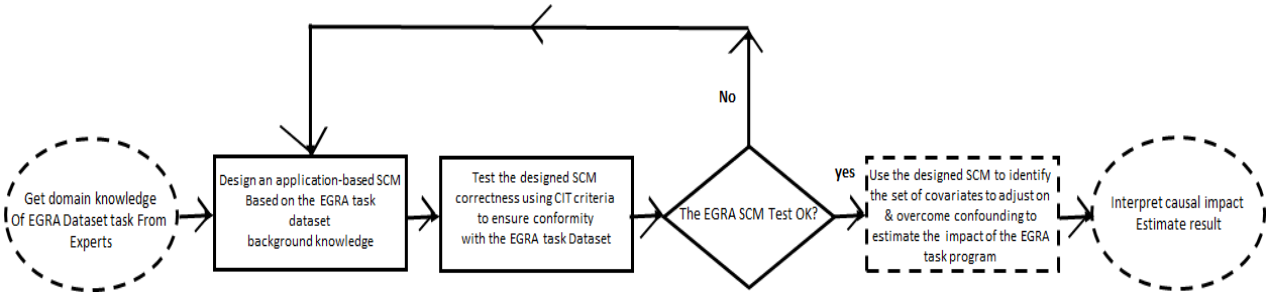
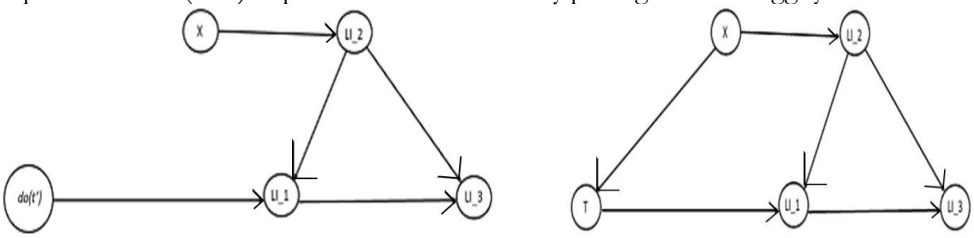


Figure 3. Shows conceptual framework procedure for performing the CIT criteria and the impact estimation

The corresponding NPSEM generated from mutilated DAG (intervention graph) as shown in Figure 4(b) for our SENS-EGRA SCM designating an intervention distribution is given as Equation (6).

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

Notice that LI_1 is not conditioned on, since from the DAG, it is considered a post-treatment or mediator variable. Pearl et al. [7, 26, 45, 55], 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 library package called Daggity of reference [56].



(a) (b)

Figure 4: Shows our SENSE-EGRA SCM with (**a**) and without (**b**) intervention

6. Results Presentation & Discussion

In this section, we present the result for the validation of the SENSE-EGRA SCM using the CIT criteria.

6.1 Results Presentation for CIT Criteria for SENSE-EGRA SCM

Designing a SCM is a qualitative process that is subjective based on background knowledge. Hence, to ensure its correctness experts advise validation and testing of the model with the dataset, which is an objective process [4, 7, 45, 57-60]. One of the most pervasive objective validation tests for SCM is the use of conditional independence testing (CIT) criteria [7, 45, 57-60]. Thus, once the validation process is over, and SCM is affirm to be the true

representation of the dataset, the adjustment criteria can be applied to the SCM. Pearl et al.[7, 45, 58], 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 (mathematics formula) for adjusting covariates and estimating the causal impact of the intervention. For the experiment in this study, we implemented the CIT using the identified conditional independencies set of Equation (4) and applied the back-door adjustment criteria for eliminating confounding bias as shown in Equations (5) and (6) respectively. Table 4 shows the results of the CIT performed on the dataset to verify and validate the correctness of our SENSE- EGRA SCM. The process is implemented in the R library package tool of reference [56], and Algorithm 1 show the CIT procedure for any EGRA SCM intervention program with binary treatment task of letter identification and Algorithm 2 shows the SENSE-EGRA SCM CIT criteria using the same task of letter identification.

Algorithm 1. Computation of CIT Criteria for any EGRA SCM with Binary Treatment for the task of Letter Identification

INPUT: X, T, L, Y
OUTPUT: RMSEA, p.value, CI
1: Start
2: Declare {X: = Set of covariates. Where X ∈ $\{x_i, \ldots x_n\}$
T: = Treatment variable. Where T∈{0,1}
Y: = Outcome variable. Where Y is continuous or categorical
L: = Other assessment or evaluation criteria variables. Where L ∈ {LI_1, LI_2}
CI: = Confidence Interval @ 95%
3: Read X, T, L, Y
4: for X: = x_1
compute {P (EGRA-SCM CIT parameters derived from background knowledge of X, T, L, & Y)}
print RMSEA, p.value, 95%CI
plot (print)
5: if RMSEA<=0, p.value <= 0.05, AND plot (print) intersects = 0 OR + 0.1 then
print "CIT validation confirmed"
else
print "CIT validation not confirmed"
6: for X:= $x_2, \ldots x_n$
Repeat steps 3-5.
7: End

Algorithm 2. Computation of CIT Criteria for SENSE-EGRA SCM for Letter Identification Task

INPUT: X, T, L, Y
OUTPUT: RMSEA, p.value, CI
1: Start
2: Declare {X: = Set of covariates. Where X ∈ {state,..., Q10}
T: = Treatment variable. Where T ∈ {0,1}
Y: = Outcome variable. Where Y: = LI_3 is continuous
L: = Other assessment or evaluation criteria variables.
Where L∈ {LI_1, &LI_2}}
CI: = Confidence Interval @ 95%

3: Read L_1, L_2, L_3, T, X
4: for X: = State
compute {P($LI\_2 \perp T|X$,
$LI\_3(Y) \perp T|LI\_1, LI\_2$,
$LI\_3(Y) \perp X|LI\_2, T$,

$LI\_3(Y) \perp X|LI\_1, LI\_2)\}$
print RMSEA, p.value, 95% CI
plot (print)
5: if RMSEA<=0, p.value <= 0.05, AND plot (print) intersects = 0 OR + 0.1 then
print "CIT validation confirmed"
else
print "CIT validation not confirmed"
6: 6. for X: = {LGA, School, Gender, Age, Q4, Q5, Q6_0,
Q6_1, Q6_2, Q6_3, Q7, Q8, Q9, Q10}
Repeat steps 3-5
7: End

Table 4. Shows the result of the CIT identified in Equation (4) for each instance of X

X		localTests for X		95% Confidence Interval	
	CIT Criteria	**RMSEA**	**p.value**	**2.5%**	**97%**
State	LI_1 ⊥ State|LI_2, T	1.934137e-02	0.23457122	0.0000000	0.9779508
	LI_2 ⊥ T|State	0.04312486	1.409349e-18	0.05049194	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	0.33122132	2.412407e-10	0.09916436	1.1437016
	LGA ⊥ LI_3 |LI_1, LI_2	0.53076862	1.184351e-01	0.00000000	1.8904874
	LGA ⊥ LI_3|LI_2, T	0.33223041	1.199982e-09	0.09886772	1.1605461
	LI_2 ⊥ T |LGA	0.08186839	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	0.1189509	1.1149223
	LI_2 ⊥ T |School	0.1169519	2.756094e-01	0.0000000	0.6437887
	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	0.26596571	2.558248e-10	0.0000000	1.15597679
	Gender ⊥ LI_3 |LI_1, LI_2	0.46412185	5.548376e-01	0.0000000	1.89753080
	Gender ⊥ LI_3|LI_2, T	0.27824380	1.766274e-09	0.0000000	1.1558097
	LI_2 ⊥ T |Gender	0.04870334	4.072550e-08	0.05287745	0.07544265
	LI_3 ⊥ T|LI_1, LI_2	0.26666667	4.3937728e-01	0.0000000	1.35522813
Age	Age ⊥ LI_1|LI_2, T	0.28759887	1.290115e-07	0.07086121	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	1.35522813
Q4	LI_3 ⊥ Q4|LI_2, T	0.22919113	2.404441e-01	0.01132237	0.9550414
	LI_3 ⊥ Q4 |LI_1, LI_2	0.06773323	7.879640e-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	1.295997e-03	0.01574901	1.05807555
	LI_1 ⊥ Q5 |LI_2, T	0.0481797	8.820992e-11	0.05611486	0.08759465
	LI_3 ⊥ Q5|LI_2, T	0.2706585	1.056003e-03	0.01555926	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	1.35522813
Q6_0	LI_1 ⊥ Q6_0|LI_2, T	0.21334892	9.796304e-02	0.007870615	0.7946580
	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	0.000000000	0.0000000

X		localTests for X		95% Confidence Interval	
	CIT Criteria	RMSEA	p.value	2.5%	97%
	LI_2 ⊥ T|Q6_0	0.26666667	4.3937728e-01	0.000000000	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	6.145748e-15	0.052218641	0.07923409
	LI_3 ⊥ Q6_1|LI_2, T	0.22164386	1.788882e-02	0.002980019	1.00789748
	LI_3 ⊥ T |LI_1, LI_2	0.27190319	4.971093e-01	0.000000000	1.50060436
	LI_2 ⊥ T|Q6_1	0.26666667	4.3937728e-01	0.000000000	1.35522813
Q6_2	LI_1 ⊥ 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	0.000000000	1.74885066
	LI_2 ⊥ T|Q6_2	0.26666667	4.3937728e-01	0.000000000	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	0.000000000	1.4658619
	LI_2 ⊥ T|Q8	0.26666667	4.3937728e-01	0.000000000	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	1.419695e-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	0.000000000	1.9380009
	LI_2 ⊥ T|Q9	0.26666667	4.3937728e-01	0.000000000	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	0.000000000	2.0001680
	LI_2 ⊥ T|Q10	0.26666667	4.3937728e-01	0.000000000	1.3552281

6.2 CIT Results Discussion

When testing for conditional independence between two or more variables, it is required that their conditional dependency be zero [56]. Hence, with the use of the R tool of reference [56], 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 designed SCM. 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 [3]. The algorithms process for performing this CIT validation for any EGRA SCM and our SENSE-EGRA SCM in the area of letter identification task are shown in Algorithms 1 and 2.

Thus the R tool produced by reference [56], has the functions LocalTests()and the PlotLocalTestResults(), which are used for the analysis of the CIT criteria. The function LocalTests()test the CIT-identified criteria for each of the feature variables instance of X (i.e., $X = State, LGA, Gender, Age,$ etc.,) under the five conditional independence conditions identified in our SENSE- EGRA SCM of Equation (4), at a confidence interval of 95% for

all test cases as shown in column 2 (i.e., the label written as 2.5% & the 97%) of Table 4. While the PlotLocalTestResults() function plots the results of the localTests() function as shown in Figure 5. All the results indicate negative p-values and zero-scale RMSEA values. Thus, validating the correctness of our SENSE-EGRA SCM, as no conditional dependency exceeds 0.4 in all test cases, as shown in PlotLocalTestResults() graphical output of Figure 5; meaning their dependence is nearly zero. Thus, confirming and validating the correctness of our SENSE-EGRA SCM as shown in Table 4.

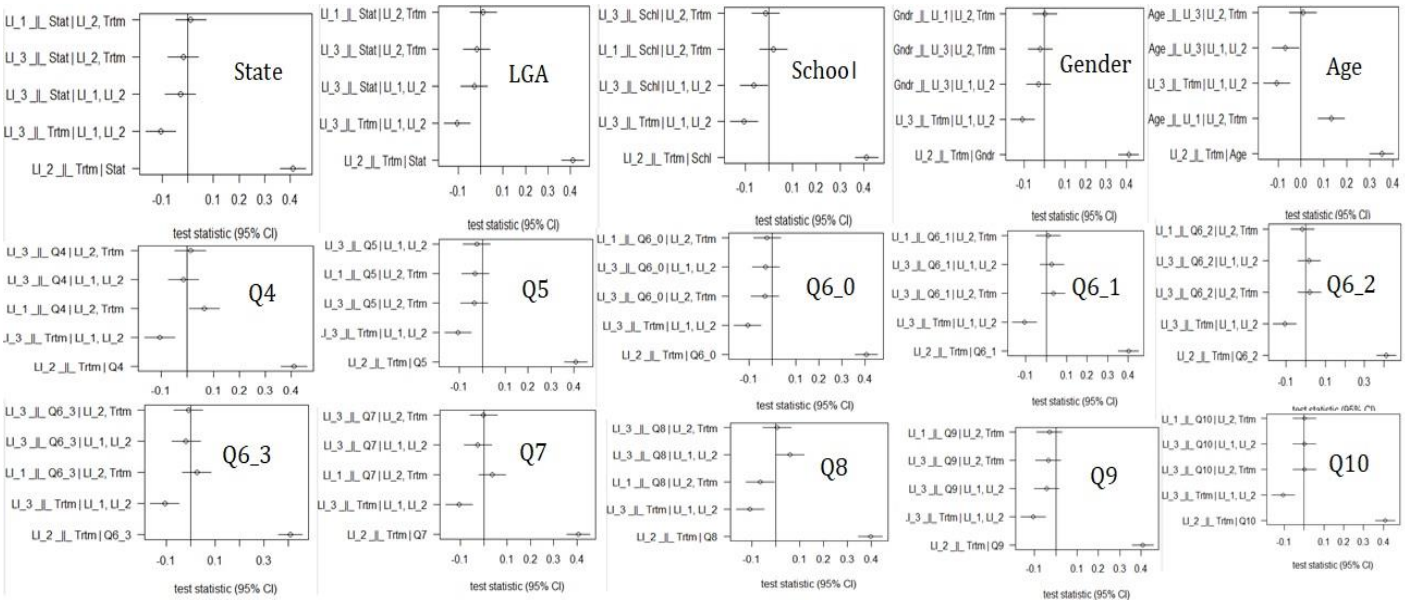


Figure 5. Shows the graph result of the CIT criteria that plots the localTest results of Table 4 for each X instance

7. Conclusion and Future Work

7.1 Conclusion

In this study, we have designed a novel application-based SCM from the background knowledge obtained from the American University of Nigeria (AUN), Yola's project on the letter identification subtask of Early Grade Reading Assessment program tagged "Strengthen Education in Northeast Nigeria - SENSE-EGRA" which was sponsored by the United States Agency for International Development (USAID), which occurred between 2021 to 2202. We employed the conditional independence test (CIT) criteria for the testing and validating of the model's '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 and the development and correctness validation testing of an application-based novel SCM for the SENSE-EGRA dataset, which could be used for the estimation of the causal impact of the intervention program under review.

7.2 Future Work

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 and propensity score by (weighting, stratification, and matching) to deal with confounding and selection biases to estimate the causal inference of the SENSE-EGRA intervention program of the American University of Nigeria, Yola, Adamawa State, Nigeria under the sponsorship of USAID.

Author Contributions: Author 1 initiated the project, performed the experiment and wrote the manuscript. Authors 2 and 3 supervised the project, validated the experiment, edited and proofread the work.

Funding: This work has no funding

Data Availability Statement: The dataset, its variable meanings, and the CIT criteria codes alongside the entire materials needed for the reproduction of this study can be accessed on our GitHub page at: https://github.com/Sadaju-Codes/SENSE-EGRA_Project.git.

Acknowledgments: We acknowledged the American University of Nigeria, Yola Adamawa State, Nigeria for making available their SENSE-EGRA intervention program dataset for this study.

Conflict of Interest Declaration: All authors have no financial or proprietary interests in any material discussed in this work.

References

[1] J. F. Box, "RA Fisher, the Life of a Scientist," *Revue Philosophique de la France Et de l*, vol. 170, no. 4, 1980.
[2] K. Benson and A. J. Hartz, "A Comparison of Observational Studies and Randomized, Controlled Trials," *N. 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," *Am. J. Med.*, vol. 122, no. 2, pp. 114–120, Feb. 2009, doi: 10.1016/j.amjmed.2008.09.030.
[4] P. W. G. 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, May 2021, doi: 10.1093/ije/dyaa213.
[5] S. Greenland, J. Pearl, and J. M. Robins, "Causal diagrams for epidemiologic research," *Epidemiology*, pp. 37-48, 1999, doi: 10.1097/00001648-199901000-00008.
[6] J. Pearl, *Causality: Models, Reasoning and Inference*, 2nd ed. University of California, Los Angeles: Cambridge University Press, 2009.
[7] J. Tian and J. Pearl, "Department of Statistics Papers A General Identification Condition for Causal Effects," *UCLA Dep. Stat. Pap.*, 2002.
[8] G. T. Ayem, S. G. Thandekkattu, A. S. Nsang, and M. Fonkam, "Structural Causal Model Design and Causal Impact Analysis: A Case of SENSE-EGRA Dataset," 2023, pp. 39–52. doi: 10.1007/978-981-99-3485-0_4.
[9] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection." *International Joint Conference on Artifical Intelligence, (IJCAI)*, pp. 1137-1145, 1995.
[10] S. Borra and A. Di Ciaccio, "Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods," *Comput. Stat. Data Anal.*, vol. 54, no. 12, pp. 2976–2989, Dec. 2010, doi: 10.1016/j.csda.2010.03.004.
[11] A. Shojaie, and E. B. Fox, "Granger causality: A review and recent advances," *Annual Review of Statistics and Its Application*, vol. 9, pp. 289-319, 2022, doi: 10.1146/annurev-statistics-040120-010930.
[12] S. L. Bressler, and A. K. Seth, "Wiener–Granger causality: a well-established methodology," *Neuroimage*, vol. 58, no. 2, pp. 323-329, 2011, doi: 10.1016/j.neuroimage.2010.02.059.
[13] E. Bareinboim, and J. Pearl, "Controlling selection bias in causal inference." *Artificial intelligence and Statistics*.PMLR, pp. 100-108, 2012.
[14] E. Bareinboim, J. Tian, and J. Pearl, "Recovering from Selection Bias in Causal and Statistical Inference (Supplemental Material)," 2014.
[15] G. T. Ayem, A. Ajibesin, A. Iorliam, and A. S. Nsang, "A mixed framework for causal impact analysis under confounding and selection biases: a focus on Egra dataset," *International Journal of Information Technology*, 2023, doi:10.1007/s41870-023-01490-6
[16] B. Piper, S. Dryden-Peterson, V. Chopra, C. Reddick, and A. Oyanga, "Are refugee children learning? Early grade literacy in a refugee camp in Kenya," *Journal on Education in Emergencies*, vol.5 (2): pp.71-107 2020, doi:10.33682/f1wr-yk6y.
[17] S. Taylor, J. Cilliers, C. Prinsloo, B. Fleisch, and V. Reddy, "The Early Grade Reading Study: Impact evaluation after two years of interventions," *EGRS Evaluation Report*, 2017.
[18] B. Fleisch, V. Schöer, G. Roberts, and A. Thornton, "System-wide improvement of early-grade mathematics: New evidence from the Gauteng Primary Language and Mathematics Strategy," *International Journal of Educational Development*, vol. 49, pp. 157-174, 2016, doi: 10.1016/j.ijedudev.2016.02.0.
[19] L. O. Costa, and M. Carnoy, "The effectiveness of an early-grade literacy intervention on the cognitive achievement of Brazilian students," *Educational Evaluation and Policy Analysis*, vol. 37, no. 4, pp. 567-590, 2015, doi: 10.3102/0162373715571437.
[20] B. Piper, and M. Korda, "EGRA Plus: Liberia. Program Evaluation Report," *RTI International*, 2011.
[21] M. Davidson, M. Korda, and O. W. Collins, "Teachers' use of EGRA for continuous assessment: the case of EGRA Plus: Liberia," *The Early Grade Reading Assessment*, pp. 113, 2011.
[22] M. Davidson and J. Hobbs, "Delivering reading intervention to the poorest children: The case of Liberia and EGRA-Plus, a primary grade reading assessment and intervention," *Int. J. Educ. Dev.*, vol. 33, no. 3, pp. 283–293, May 2013, doi: 10.1016/j.ijedudev.2012.09.005.
[23] M. M. de Oca, J. Hill, L. Aber, C. T. Dolan, and K. Gjicali, "The Impact of Attending a Remedial Support Program on Syrian Children's Reading Skills: Using BART for Causal Inference," *arXiv preprint* arXiv:2208.13906, 2022.
[24] R. Guo, L. Cheng, J. Li, P. R. Hahn, and H. Liu, "A Survey of Learning Causality with Data," *ACM Comput. Surv.*, vol. 53, no. 4, pp. 1–37, Jul. 2021, doi: 10.1145/3397269.
[25] C. Hitchcock, and M. Rédei, "Reichenbach's common cause principle," 2020, online at.< Reichenbach's Common Cause Principle (Stanford Encyclopedia of Philosophy>.
[26] J. Pearl, *Causality*. Cambridge University Press, 2009. doi: 10.1017/CBO9780511803161.
[27] J. Peters, D. Janzing, and B. Scholkopf, *Elements of causal inference*. London, England: MIT Press, 2017.

[28] J. Neyman, "Sur les applications de la theorie des probabilities aux experiences agricoles: essai des principes (Masters Thesis); Justification of applications of the calculus of probabilities to the solutions of certain questions in agricultural experimentation. Excerpts English translation (Reprinted)," *Stat Sci*, vol. 5, pp. 463-472, 1923.
[29] D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, vol. 66, no. 5, pp. 688, 1974.
[30] V. Karwa, A. B. Slavković, and E. T. Donnell, "Causal inference in transportation safety studies: Comparison of potential outcomes and causal diagrams," *The Annals of Applied Statistics*, vol 5. 2011, doi: :10.1214/10-AOAS440.
[31] C. R. Lesko, A. L. Buchanan, D. Westreich, J. K. Edwards, M. G. Hudgens, and S. R. Cole, "Generalizing Study Results," *Epidemiology*, vol. 28, no. 4, pp. 553–561, Jul. 2017, doi: 10.1097/EDE.0000000000000664..
[32] G. W. Imbens, "Potential outcome and directed acyclic graph approaches to causality: Relevance for empirical practice in economics," *Journal of Economic Literature*, vol. 58, no. 4, pp. 1129-1179, 2020, doi: 10.1257/jel.20191597.
[33] P. Spirtes, C. Glymour, and R. Scheines, "Discovery Algorithms for Causally Sufficient Structures," 1993, pp. 103–162. doi: 10.1007/978-1-4612-2748-9_5.
[34] S. Lauritzen, "Causal Inference from Graphical Models," 2000. doi: 10.1201/9781420035988.ch2.
[35] F. Eberhardt, "Introduction to the foundations of causal discovery," *International Journal of Data Science and Analytics*, vol. 3, no. 2, pp. 81-91, 2017, doi: 10.1007/s41060-016-0038-6.
[36] J. Y. Halpern, "The Book of Why, Judea Pearl, Basic Books (2018)," *Elsevier*, 2019.
[37] B. Neal, "Introduction to causal inference from a machine learning perspective," *Course Lecture Notes* (draft), 2020.
[38] F. Elwert, "Graphical Causal Models," 2013, pp. 245–273. doi: 10.1007/978-94-007-6094-3_13..
[39] A. R. Nogueira, A. Pugnana, S. Ruggieri, D. Pedreschi, and J. Gama, "Methods and tools for causal discovery and causal inference," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, pp. e1449, 2022, doi: 10.1002/widm.1449.
[40] L. Yao, Z. Chu, S. Li, Y. Li, J. Gao, and A. Zhang, "A survey on causal inference," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 15, no. 5, pp. 1-46, 2021, doi:10.1145/3444944.
[41] C. Glymour, K. Zhang, and P. Spirtes, "Review of causal discovery methods based on graphical models," *Frontiers in genetics*, vol. 10, pp. 524, 2019, doi: 10.3389/fgene.2019.00524.
[42] J. Pearl, Probabilistic reasoning in intelligent systems: networks of plausible inference: *Morgan Kaufmann*, 1988.
[43] L. Gultchin, M. Kusner, V. Kanade, and R. Silva, "Differentiable causal backdoor discovery." in *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, PMLR, pp. 3970-3979, 2020, doi:108:3970-3979.
[44] J. Correa, and E. Bareinboim, "Causal effect identification by adjustment under confounding and selection biases."in *Thirty-First Conference on Artificial intelligence*, San Francisco, CA, 2017.
[45] J. Pearl, and D. Mackenzie, The book of why: the new science of cause and effect: Basic books, 2018.
[46] J. Pearl, "Theoretical impediments to machine learning with seven sparks from the causal revolution," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining - WSDM*, 2018, doi: 10.1145/3159652.3176182.
[47] T. S. Richardson, and J. M. Robins, "Single world intervention graphs: a primer.", 2013.
[48] J. Zhang, and P. Spirtes, "Intervention, determinism, and the causal minimality condition," *Synthese*, vol. 182, no. 3, pp. 335-347, 2011, doi: 10.1007/s11229-010-9751-1.
[49] A. Hauser and P. Bühlmann, "Characterization and greedy learning of interventional markov equivalence classes of directed acyclic graphs," *J. Mach. Learn. Res.*, vol. 13, pp. 2409–2464, 2012.
[50] A. Hauser, and P. Bühlmann, "Jointly interventional and observational data: estimation of interventional Markov equivalence classes of directed acyclic graphs," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 77, no. 1, pp. 291-318, 2015, doi:10.1111/rssb.12071.
[51] R. Mayrhofer, and M. R. Waldmann, "Sufficiency and necessity assumptions in causal structure induction," *Cognitive science*, vol. 40, no. 8, pp. 2137-2150, 2016, doi:10.1111/cogs.12318.
[52] J. Zhang, and W. Mayer, "Weakening faithfulness: some heuristic causal discovery algorithms," *International Journal of data science and Analytics*, vol. 3, no. 2, pp. 93-104, 2017, doi: 10.1007/s41060-016-0033-y.
[53] J. Zhang, and P. L. Spirtes, "Strong faithfulness and uniform consistency in causal inference," in *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence (UAI2003)*, 2012, doi:10.48550/arXiv.1212.2506.
[54] American University of Nigeria (AUN), "SENSE project evaluation: reading proficiency assessment, results of the hausa early grade reading assessment (EGRA) in Adamawa And Gombe," a report, 2021.online at.< SENSE-EGRA_Project/SENSE_EGRA Evaluation report_20211026.pdf at main · Sadaju-Codes/SENSE-EGRA_Project (github.com)>.
[55] I. Shpitser, T. VanderWeele, and J. M. Robins, "On the validity of covariate adjustment for estimating causal effects," *Proc. 26th Conf. Uncertain. Artif. Intell. UAI 2010*, pp. 527–536, 2010.
[56] A. Ankan, I. M. Wortel, and J. Textor, "Testing graphical causal models using the R package "dagitty"," *Current Protocols*, vol. 1, no. 2, pp. e45, 2021, doi:10.1002/cpz1.45
[57] F. Thoemmes, Y. Rosseel, and J. Textor, "Local fit evaluation of structural equation models using graphical criteria," *Psychological methods*, vol. 23, no. 1, pp. 27, 2018, doi:10.1037/met0000147.
[58] J. Pearl and T. S. Verma, "A theory of inferred causation," 1995, pp. 789–811. doi: 10.1016/S0049-237X(06)80074-1..
[59] J. Pearl and E. Bareinboim, "Transportability of Causal and Statistical Relations: A Formal Approach," *Proc. AAAI Conf. Artif. Intell.*, vol. 25, no. 1, pp. 247–254, Aug. 2011, doi: 10.1609/aaai.v25i1.7861.
[60] J. Pearl, "Causal inference in statistics: An overview," *Statistics surveys*, vol. 3, pp. 96-146, 2009, doi: 10.1214/09-SS057.