ESP-JETA
ESP Journal of Engineering & Technology Advancements
ISSN: 2583-2646 / Volume 3 Issue 4 November 2023 / Page No: 1-22
Paper Id: JETA-V3I8P101 / Doi: 10.56472/25832646/JETA-V3I8P101

Research Article

Causal Inference Estimates with Backdoor Adjustment Condition vs. the Unconfoundedness Assumption: A Comparative Analysis Study of the Structural Causal Model and the Potential Outcome Frameworks

Gabriel Terna Ayem[1], Ozcan Asilkan[2], Aamo Iorliam[3], Rabiu Ibrahim[4], Salu George Thandekkattu[5]

[1,2,5]*Computer Science Department, School of Information Technology and Computing, American University of Nigeria, Yola, Nigeria.*
[3]*Data Science Department, School of Information Technology and Computing, American University of Nigeria, Yola, Nigeria*
[4]*Software Engineering Department, School of Information Technology and Computing, American University of Nigeria, Yola, Nigeria*

Corresponding Author: Gabriel Terna Ayem

Received Date: 10 September 2023 Revised Date: 23 September 2023 Accepted Date: 01 November 2023

Abstract: *We present an empirical and theoretical comparative analysis of the structural causal model (SCM) and the potential outcome (PO) frameworks under the presence of biases of confounding and selection in the dataset. We used the Early Grade Reading Assessment (EGRA) dataset tagged "Strengthening Education in Northeast (SENSE) Nigeria, - an educational intervention program of the American University of Nigeria under the sponsorship of the United States Agency for International Development (USAID). The SCM backdoor adjustment criteria and the unconfoundedness assumption of the PO framework and other assumptions are employed to overcome confounding bias in the dataset to estimate the causal inference of the SENSE-EGRA intervention program. Further, we employed 4 statistical estimators for the adjustments of covariates and the estimation of the causal inference for the intervention program. These estimators are structured to handle selection bias during covariates adjustment, by carefully matching up the treated and control units in the dataset. The 4 estimators are. Viz. The ordinary least square regression adjustment (OLS), the propensity score weighting (PSW), the propensity score stratification (PSS), and the propensity score matching (PSM) adjustments. The results evinced higher average treatment effects (ATE) estimates for the SCM framework than that of the PO's ATE estimates, which are (SCM'). Viz. OLS = 0.661, the PSS = 1.592, the PSM = 2.173, and the PSW = 4.931, while the PO is. Viz. The PSW = 0.035, the PSS = 0.036, the OLS = 0.079, and the PSM = 0.066. These disparities in the ATE results for the two frameworks are due to the fact, that the covariates adjustment was applied to all the covariates in the dataset including the mediator variable under the PO framework, which is considered a forbidden act under the SCM framework. Additionally, the explicit representation of the dataset generation process in a direct acyclic graph (DAG), and the concomitant proving of the same, clarifies and validates the inference estimates from the SCM framework - a key component that is lacking in the PO framework.*

Keywords: *Backdoor Adjustment, Comparative Research, Potential Outcome, Structural Causal Model, Unconfoundedness Assumption.*

I. INTRODUCTION

Randomized control trial (RCT) experiments also known as A/B tests, since their inception till today remain the gold standard for analyzing and estimating causal claims. This is so because A/B tests are exceptionally good in de-confounding variables (other variables in the datasets aside from the treatment and outcome variables) that may confound or make the treated and controlled unit variables (collectively called the treatment variable) difficult to match up during the estimation of the ATE. The ATE estimation is the process of determining the effects of the treatment variable over the outcome variable. Despite the A/B test's superb performance in dealing with confounding and effectively estimating the phenomenon of cause and effects in an experiment, some causal phenomena scenarios make the A/B test simply infeasible, unethical, or too expensive to perform. Take for instance, an A/B test that seeks to determine the effects of smoking on health and well-being, the A/B experimental setup must involve setting aside a group of people called treated who will be constantly exposed to smoking within a given period and then setting aside a second group of people, called the controlled, who will not be exposed to smoking. After which the experimenter would seek to determine the causal effect of the treated (those who smoked), and the controlled (those who didn't smoke) on their health and well-being. This scenario paints a picture of an RCT experiment that is simply unethical to perform. In the same vein, similar A/B tests that are too expensive and practically infeasible to perform exist. Observational datasets are however common, even with causal scenarios where the A/B test is impossible, you could have a dataset on people who smoke on their own and those who have made the choice not to smoke. Thus, with these kinds of datasets available, researchers in recent times have been able to uncover ways of using these sets of datasets to simulate an A/B test, once the confounding variables, which make the two groups incomparable are effectively dealt with [1-5]. One such method known as the SCM framework involves making some qualitative causal assumptions on the process of the data generation and then encoding the assumptions in direct acyclic graphs (DAGs). The DAGs are represented with nonparametric structural equations (NPSEM) that obey the Markov condition (MC) principles, and then later used with any parametric estimator for the estimation of the causal inference for the dataset whose causal questions are posed. Thus, the DAGs enable the identification of the confounding variables that are likely to influence the causal estimation of the treatment on the outcome variable and are consequently removed or de-confounded via the simulation of an A/B test process known as the do-calculus intervention, and some adjustment technique called the backdoor adjustment. Thus, Pearl [4] who proposed this framework defines confounding as the difference between the probability distribution and the do-calculus intervention distribution. The do-calculus intervention is written as $(P(y|x) = P(y|do(X = x)))$, and it connotes setting one of the units of the treatment variable (treated or controlled) to a fixed or constant value. Further, a method called the conditional independent test (CIT) criteria enables the testing and validation of the assumptions encoded in the DAG under the SCM framework. Overall, once the SCM components (DAGs, and NPSEMs) are replete and correct, the issue of confounding bias can be dealt with through identification and relevant adjustment [4, 6], making the ATE or the average treatment effects for the treated (ATT) estimates from the model to be trusted or to reflect the causal claims in such a dataset.

Similarly, another method for causal analysis and estimation proposed by Donald Rubin called the Potential outcome (PO) framework, also known as the Rubin Causal Model (RCM) [7-11] can be used to analyze and estimate causality in observational datasets. Although this method does not explicitly represent causal assumptions in a DAG but uses a table to represent actual and potential outcomes for the observed and unobserved (counterfactual) causal phenomenon of interest (which is the treated and controlled units and other variables associated with them). Further, It employs assumptions such as the stable unit treatment value assumption (SUVTA) [8, 10, 12, 13], the unconfoundedness [14] or ignorability [15] assumption, which can be likened to the backdoor adjustment criteria of the SCM framework. Also, another assumption called the overlap [16] assumption is used to handle selection and even confounding biases within this framework. With these sets of assumptions, it is possible to estimate ATT and ATE in observational datasets that seek to solve causal questions. Hence, the main objective of this study is to explore the two frameworks with particular reference to the backdoor adjustment criteria of the SCM and the unconfoundedness assumption of the PO framework with the view to estimating ATE and ATT using the SENSE-EGRA intervention dataset since the dataset contains a set of confounding variables. Also, aside from the empirical estimation of causal inference for the said dataset, which results are used for the comparison of the two frameworks, the literature review section shall establish and explicate their theoretical foundation and differences as well.

A few studies on the comparative analysis of SCM and PO frameworks exist in the extant literature. Aliprantis et al [9] made a theoretical distinction between the causal effects of an SCM and Rubin causal model (PO) framework. Albeit no empirical contribution involving the estimation of causal inference between the two frameworks was made. Vishesh et al [17], in their work employed causal diagrams (DAG), which is a component of an SCM and the PO framework in the estimation and comparison of causal inference in a specific transportation safety problem. That is, they sought to estimate the causal effects of pavement marking retro-reflectivity on the safety of road segments, using real-life datasets in the area. They employed three main estimators in the estimation of the ATE, - regression adjustment, PSW, and causal Bayesian networks (CBN). Their study which was both theoretical and empirical specifically applied and contributed to transportation space. Markus [18], made a generic and comprehensive theoretical comparison of causal inference estimation with the two frameworks of PO and SCM and further encouraged more comparative research that would help develop each of the frameworks and bring about an in-depth understanding of their relative strength and weaknesses. Imbens [19] in his comparative essay on the topic, discussed the PO and SCM approach to causality and their relevance for empirical works in economics, focusing on casual questions that are best answered by each of the frameworks. Further, they explained in some detail why much of the field of economic research applications is more suited to the PO framework than the SCM's.

A. Background of Study and Comparison with Similarly Related Works:

The SENSE-EGRA educational intervention program of the American University of Nigeria (AUN), Yola, Adamawa State under the sponsorship of the USAID was aimed at improving the literacy of over 200,000 primary school pupils targeting two northeastern states of Adamawa and Gombe in Nigeria. These States were considered to be the most hit by the dreaded Boko Haram terrorist group in the region (Northeastern states), making it have the highest number of out-of-school children [20]. The aftermath analysis and impact evaluation of the SENSE-EGRA intervention program using statistical tools was found to have a positive impact on all subtasks performed, with the inclusion of the task of letter identification. Albeit causal claims could not be elicited from the statistical evaluation process, since the program was not designed with the setting of an A/B test. Thus, an objective of this work was to employ the SCM and the PO frameworks alongside the four estimators to estimate the causal impact of the SENSE-EGRA intervention program. The combinations of the SCM and the PO using these four estimators can deal optimally with the issues of confounding and selection biases and enable us to perform this causal analysis of the SENSE-EGRA intervention program on the subtask of letter identification, without biases [21]. The implementation of these two frameworks (SCM & PO) for causal impact analysis on the EGRA datasets is novel as far as we can check. See Tables 1 and 2 for the novelty of our work in comparison with previous similar works.

Other similar EGRA studies have implemented different approaches for estimating the intervention impact of different EGRA intervention programs as they relate to letter identification tasks amongst others in different parts of the world as shown in these references [22], [23], [24], [25], [26-28]. See Table 1 for the summary of the methods employed and the country or places of interventions. Albeit, a study by de Oca, Hill, Aber, Dolan and Gjicali [29] implemented the PO framework and the machine learning (ML) predictive model called the Bayesian Additive Regression Tree (BART) as an estimator for dealing with confounding and selection biases to estimate the causal impact of the Arabic-EGRA intervention program task on letter identification and other subtasks. Thus, our study uses the combination of the SCM and the PO frameworks with the four estimators (OLS, PSS, PWS, and PSM) instead of the BART ML model to overcome confounding and selection biases to estimate the causal impact of our focused SENSE-EGRA intervention task on letter identification for grade 2 students. Table 2 evinced the major differences between our study and theirs.

B. Study Objectives & Contributions:

The following narrative explicates the main objectives and contributions of this study.

a) Study Objectives:

The following objectives are elicited from this study:

(i) To analyze and explicate the SCM and the PO frameworks' theoretical foundation for non-experimental causation and its techniques for handling confounding bias along the four estimators employed to deal with selection bias in datasets.

(ii) To perform a causal impact analysis and estimation experiment on the SENSE-EGRA intervention program dataset on the task of letter identification for grade 2 students.

b) Study Contributions:

The following research contributions are evinced from this study.

(i) *Conceptual framework Design:* We designed a conceptual framework guideline for the proposed method for analyzing and estimating causal impact for the EGRA observational intervention datasets with confounding and selection biases.

(ii) *Model Design:* We designed and implemented for the first time a novel application-based SCM that describes the causal structure of the SENSE- EGRA dataset in the area of the letter identification task.

(iii) *Algorithm Design:* We designed a general and specific algorithm that can be used for the EGRA SCM model's validation process.

(iv) *Model Validation:* We tested and validated our SENSE-EGRA SCM for correctness using the CIT criteria implemented in a state-of-the-art tool of reference [30].

(v) *Causal Impact Evaluation:* We evaluated the causal impact estimation of the SENSE- EGRA intervention program in the targeted subtask of letter identification, using 4 different estimation techniques that optimally handle selection bias during the adjustment of the covariate set in the dataset.

(vi) *Experiment Reproducibility:* The codes and SENSE- EGRA datasets for this experiment are made available free of cost online for the reproducibility of the experiment. See the appendix for the data and codes for the experiment reproducibility link.

For the rest of the work, in section 2, we present the definitions and theoretical foundation of the methods and frameworks applied to this work, i.e., the SCM and PO frameworks alongside the four estimators, with the adjustment criteria phenomenon required to deal with confounding and selection biases. Section 3 introduces the SENSE-EGRA dataset experiment and explains how the methodology of an SCM is used to design and identify the adjustment set. Similarly, in the same section, the CIT criteria and its algorithm process that are required to test the model's correctness are explicated. In section 4 the results from the CIT carried out on the model alongside the result of the causal impact estimation of the SENSE-EGRA intervention program using the 4 estimators are presented for both SCM and PO frameworks alongside the discussion and interpretation of the results. In section 5, we discussed the differences in the results for both frameworks. In section 6, we conclude the study with a positive impact causal claim for the SENSE-EGRA intervention program and briefly peek into our future work.

Table 1: Shows Previous EGRA Studies on Letter Identification and Other Tasks and the Impact Evaluation Methods Employed

Paper	Area(s) of EGRA Intervention/ Focused Grade Level	Impact Evaluation Method Employed	Name of EGRA Intervention/ Country
[29]	Letter recognition and 4 other tasks/Grades 1-9	Potential Outcome and Bayesian Additive Tree (BART) as Estimator	Arabic EGRA assessment: Lebanon and Syria
[22]	Letter-sound fluency and 6 others/Grades 1-3	Descriptive and Inferential statistics	The Kakuma and Kalobeyei Schools' EGRA program: Kenya
[23]	Letter sound recognition and 7 others/Grades 1-3	Randomized control Trial (RCT) design and impact evaluation with Descriptive and Inferential statistics	Improving Early Grade Reading in South Africa: South Africa
[24]	Letter identification, reading, and Mathematics /Grades 1 & 2.	Quasi-experimental design, Descriptive and Inferential statistics	Learning for Living project: South Africa
[25]	Portuguese language reading (letter identification inclusive) and mathematics/Grades 1-9	A descriptive and a triple t-test Inferential statistics comparison analysis method called Difference-in-Difference-in-Difference [DDD]	Literacy Program at the Right Age (Pacto pela Alfabetização na Idade Certa [PAIC]): Brazil
[26-28]	Letter naming fluency and 6 others/ Grades 2 & 3	Randomized control Trial (RCT) design and impact evaluation with Descriptive and Inferential statistics	EGRA Plus: Liberia

Table 2: Shows the Differences between the Study by de Oca, Hill, Aber, Dolan and Gjicali [29] and ours as it Relates to Methods, Assumptions, and Model Employed

Differences Between de Oca, Hill, Aber, Dolan and Gjicali [29], and Our Study		
COMPARISON INDICES	**de Oca, Hill, Aber, Dolan and Gjicali [29] STUDY**	**OUR STUDY**
Causal Framework employed	PO Framework.	PO & SCM Frameworks.
Assumptions used in overcoming confounding bias in covariates set	Unconfoundedness (ignorability), stable unit treatment value assumption (SUTVA), & the Overlap assumption.	All PO assumptions plus the 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 the possibility of adjusting on a 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 from the process. See Section 3, Fig. 4, 5 & and 6, and Table 3 for model design, adjustment criteria identification, algorithms process, and model validation via CIT criteria.
Estimator(s) used for covariates set adjustment/balancing & causal impact estimation (selection bias)	BART	OLS, PSW, PSS, & PSM

II. DEFINITION AND THEORETICAL FOUNDATIONS TO THE STUDY

Causal impact analysis with observational datasets under the presence of confounding and selection biases can best be articulated with two formal frameworks of SCM [2, 4] and PO [7, 8, 10, 11]. These two formal languages are the basis for the characterization and assumptions underlying the data-generating process of the dataset. These fundamental frameworks are consistent in simulating an A/B test when it comes to the representation, analysis, and estimation of causality with observational datasets that have confounding and selection biases in them.

In this paper, just like is common with papers in the field, capital letters such as X represent a variable set. While their lower case 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 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.

A. SCM Theoretical Background:

SCM model enhances the articulation of causal theory in a given observational dataset [11]. SCM models are made up of two components. Viz. (i) DAGs and (ii) Non-Parametric structural equations (NPSEM). The DAGs are factorized based on the Bayesian network factorization with respect to the Markov condition (MC), with arrowheads showing the causal relations amongst variables. And no arrows showing the conditional independence among variables. The variables formation of a DAG may include, the treatment variable, the outcome variable, and sometimes observed and even unobserved variables. Thus, in this study, we will only consider the directed acyclic graph and not the directed cyclic graphs where the path returns to the same node.

Definition 1: Causal graphs. A graph $\mathcal{G}$ with a set of node $\mathcal{V}$ and a set of edges $\mathcal{E}$, written as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is said to be a causal graph if the set of edges $\mathcal{E}$ linking the node set V is directed based on the MC. Hence, describing the causal effects between the variables set V. Thus an edge from the treatment variable t to an outcome variable y describes a causal effect of t on y.

Definition 2: The Markov Condition (MC). The MC also called the truncated factorization (TF) [16, 17] states that a variable node is conditionally independent of its non-descendant, given its parents in a joint probability distribution. This can be expressed mathematically as shown in equation 1.

$$P(X) = \prod_{x_i \in X} P(x_i|pa(x_i)) \quad (1)$$

Where $pa(x_i)$ denotes the parent of the variable (x_i) and $i = 1 \ldots\ldots n$

Based on the MC, the estimand for the SCM model in Fig. 1 (a) factorizes as in Equation 2, instead of as in Equation 3.

$$P(X, Y, Z) = P(x)P(z|x)P(y|z) \quad (2)$$

$$P(X, Y, Z) = P(x)P(z|x)P(y|z,\ y,\ x) \quad (3)$$

Since the MC is not a sufficient assumption to empirically determine conditional dependencies (or independencies) encoded in SCM, which are the basis for confounding biases, Pearl's [1, 2] popular adjustment criteria that is hinged on, d-separation (dependency) technique is imperative to de-confound the confounders in a dataset.

Definition 3: Adjustment: A causal graph $\mathcal{G}$ with a set of node $\mathcal{V}$ with pairwise disjoint sets $T, X, Y \in \mathcal{V}$ (e.g. Fig 2) and a set of edges $\mathcal{E}$, written as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, the covariate set $X \in \mathcal{V}$ is a necessary adjustment condition for estimating the causal effect of T on Y, if for every random distribution $P(v)$ compatible with $\mathcal{G}$ it is found that the probability of y given we condition on a constant value of t, i.e., $do(t = t')$ is as shown in equation 4.

$$P(y|do(t = t')) = P(y|t, x)P(x) \quad (4)$$

Equation 4 is called an intervention distribution and in determining an adjustment set that is related to the variables T and Y, with equation 4, it is possible to identify the associated causal effects of T on Y. A lot of techniques exist for determining whether a set X is a suitable and valid adjustment, albeit the "Backdoor adjustment criteria" proposed by Pearl [1, 2] is by far the most popular one and it is defined in definition 4.

Definition 4: Adjustment (Backdoor criteria): For the backdoor adjustment criteria of a set of covariates x in a SCM to be satisfied, the following conditions must be satisfied [13], see Fig. 1.

1. x The middle node in a chain and or a fork must block all paths that are not causal from treatment t to the outcome y (this is achieved by conditioning on the node x);
2. No collider node x or its descendant must lie on the causal path from the treatment variable t to the outcome variable y.

In a SCM where the covariate adjustment criteria are sufficient and admissible, the causal estimand can be identified since the confounding bias that is occasioned by the covariate is eliminated by the adjustment criteria. Thus, the causal estimates become easy to compute. Further, the total effect of the treatment variable t on the outcome variable y is adjusted for by the set of covariates x where required [13]. The complete DAG conditions for ascertaining whether a set of covariate x is sufficient and admissible for adjustment can be derived from these papers [31-33]. Theoretically and empirically, the purpose of the adjustment criteria is to enable the right conditioning, stratifying, or blocking of a path in the SCM and then the dataset that does not convey causation and leaving open the path that does carry causation as the sole focus of the analysis and estimation. In an SCM, the chain, the fork, and the collider are identified as the three sources of confounding biases that can be found in an observational dataset, as shown in Fig. 1. Thus, the purpose of the adjustment condition is to strip off the unwanted correlation relations among the variables occasioned by the confounding bias and leave out the causal path free for causal inference estimation.

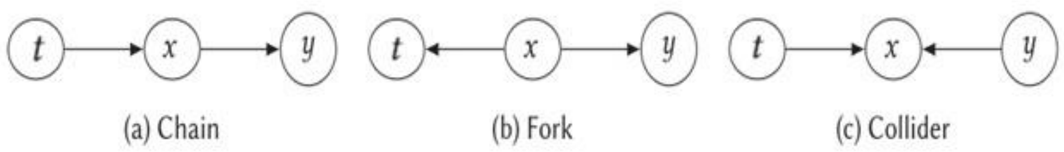


Figure 1: The Three Basic Sources of Confounding Structures in an Observational Dataset

Non-parametric Structural Equation (NPSEM): This is the second part of the SCM and it enables us to specify the causal effects depicted by the directed edges in the causal DAG. Equation 5 expresses the NPSEM of Fig. 2(a) as shown below.

$$x = \mathrm{F}_x(U_x),\ t = \mathrm{F}_t(x, U_t),\ y = \mathrm{f}_y(x,\ t,\ U_y) \quad (5)$$

In equation 5 the terms U_x, U_t, and U_y depict the noise terms for the variables observed and they are exogenous or mutually independent sources of variation that are not measured. Hence, it should be noted that the variables on the right-hand side (RHS) (usually called the endogenous variables) affect the variable on the left-hand side (LHS) and not converse since it is a matter of causation. Therefore, equation 5 provides a quantitative or parametric way of representing intervention for the corresponding causal DAG (Fig 2(a)). Intervention in SCM is defined by the use of the do-calculus, proposed by Pearl [1]. The do-calculus uses the do-operator which is expressed as $do(T = t)$, which shows the intervention of setting treatment T to some value t' as shown in Equations 6, and 7.

Definition 5. The Interventional distribution with the do-operator: The interventional distribution sometimes called post-intervention [17] is expressed mathematically as $P(y|do(t')$ which is the distribution of y when the data generation process is modified by setting the value of the variable t to t'. therefore, the SCM of Fig. 2(a) under an intervention is shown in the SCM of Fig. 2(b), and the NPSEM for the intervention of Fig. 2(b) can be expressed as shown in Equations 4, 6, and 7.

$$x = \mathrm{F}_x(U_x),\ t = t',\ y = \mathrm{F}_y(x,\ t,\ U_y) \qquad (6)$$

$$P(y|do(t') = \mathrm{F}_y(x,\ t,\ U_y) \qquad (7),$$

It should be noted that with SCM the intervention distribution, written as $P(y|do(t))$ is different from the observed probability distribution, written as $P(y|t)$ and the difference is a result of the presence of confounding which is occasioned by the covariate set X. Thus, equation 7 is the basis for simulating an A/B in datasets that ensures confounding biases are removed to carry out causal impact analysis.

Figure 2: Showing Intervention (b) and non-intervention (a) SCM

Estimating Treatment Effects. when it comes to the causal effect estimate of the treatment t on the outcome y, for the SCM in Fig. 2 for instance, the queries about the intervention distribution $P(y|do(t))$ with different values for t can be used for the estimate as shown in equation 8.

$$r(t, t') = E[y|do(t)] - E[[y|do(t')],\ t > t' \qquad (8)$$

In most causal inference problems, including ours, where the treatment t usually assumes a binary value, 0 or 1, the ATE can be expressed as the expectation of the intervention distribution, when treatment is 1 minus the expectation of the intervention distribution when treatment is 0, over the entire population as shown in Equation 9.

$$r(1,\ 0) = E[y|do(t = 1)] - E[[y|do(t = 0)] \qquad (9)$$

B. Potential Outcome (PO) Theoretical Background:

SCM framework does a great job in the area of articulating domain knowledge in a way and manner that explicitly identifies the right set of covariates that are admissible for adjustment, and thus, overcomes confounding bias, which ordinarily would not be obtained from the dataset alone. SCM queries the *do*-calculus intervention distribution to estimate ATE. Albeit, when it comes to causal inference estimation, the PO framework does a great job in handling selection bias when the strong conditional ignorability assumption is employed, with a dataset whose covariates are all measured. These strong ignorability assumptions together with the estimating techniques of propensity scores and regression adjustments, which are known for their superb matching techniques presented in section 2.3 can estimate the ATE and the conditional average treatment effects (CATE) effectively. The PO, also known as the Rubin causal model (RCM) [10, 11] framework is capable of analyzing and estimating causal effects with minimal issues. Since its introduction in 1974 by Rubin, the PO has been popularly used in the fields of social science and economics, and it can be logically equated to the SCM. Similarly to SCM, the PO model assumes random variables of T, Y $(y_i(0), y_i(1))$, and X. where T is treatment, Y is the potential outcome, which assumes two values for a binary case ($y(0)$ for controlled and $y(1)$ for treated), and X which is a set of covariates. The PO can formally be defined as explained under definition 6.

Definition 6: The Potential Outcome: Given a pair of treatment-outcome variables (t, y), the potential outcome of an individual instance i, written as $y_i(t)$, is defined as the outcome that the individual instance would have observed had the individual received treatment t [12, 31, 34-36].

Thus, with the PO framework, the underlying challenge of causal inference, which is missing data is easily articulated, as only one PO of an individual instance can be observed at a time. Hence, with the use of the PO framework, the individual Treatment effect (ITE) can be estimated as the difference between the PO of a certain instance under two different treatments (treated and controlled). This can be further extended as the average treatment effects (ATE) for a population sample. For most empirical implementations, the treatment usually assumes a binary treatment, written as $(t \in \{0,1\})$ where 1 stands for the treated instances and 0 stands for the controlled instances.

Definition 7: Individual Treatment Effects (ITE): Given an instance and its PO outcome, i and $y_i(t)$ under a binary treatment, the ITE is defined mathematically as shown in equation 10.

$$\tau_i = y_i(1) - y_i(0) \tag{10}$$

Thus, based on the ITE and the ATE, other subpopulations in the sample such as the conditional average treatment effects (CATE) can similarly be defined. Similarly, just as the do-operator for the intervention distribution that produced the ATE as explicated in Definition 5 and Equations 8 and 9, based on the SCM framework, we can also define or formulate the ATE for the PO framework as the expectation of the ITE over the entire population as shown in Equation 11.

$$\tau = E_i[\tau_i] = E_i[y_i(1) - y_i(0)] = \frac{1}{n}\sum_{i=1}^{n}(y_i(1) - y_i(0)) \tag{11}$$

Where $i = 1 \ldots\ldots n$

Thus, generating the statistical identification estimand for equation 11, with respect to the unconfoundedness assumption for a binary treatment, which can be used for the adjustments of the covariates when estimating the ATE, we take the expectation of the individual outcomes, under treated. i.e., when $= 1$ given or conditioned on the set of covariates, minus the expectation of the individual outcomes under the controlled, i.e., when $t = 0$ given or conditioned on the set of covariates x, as shown in equation 12. Thus, Equation 12 which is the ATE can be likened to the SCM intervention distributional difference which is expressed in Equation 9.

$$\tau = E_i[y_i(1) - y_i(0)] = E_X[E[\, y_i|t = 1, x] - E[\, y_i|t = 0, x] \tag{12}$$

Hence, the actual estimation and adjustment process for equation 12 can take the form of equation 13 below:

$$\frac{1}{n}\sum_x[E[\, y_i|t = 1, x] - E[\, y_i|t = 0, x]] \tag{13}$$

Thus, just like the SCM where the intervention distribution, written as $P(y|do(t))$ is different from the observed probability distribution, written as $P(y|t)$, the PO framework observed distribution, written as $P(y_i(t))$ is different from the intervention distribution, written as $P(y|t = 1))$. This is the basis for estimating the ATE which is usually the focus of causal estimate when using the afore-mentioned adjustments techniques of OLS, PSW, PSS, and PSM, which effectively matches the treated units with the controlled (untreated) units and thus, mitigating or eliminating the issue of selection bias in the dataset.

Empirically, evaluating Equations 12 and 13 all boils down to employing the mean absolute error (MAE) when the variables in the dataset are homogenous. Thus, the given truth τ in the dataset and the predicted or inferred truth $\hat{\tau}$ is used to estimate the ATE using Equation 14. In a case where the variables are heterogeneous, the CATE is estimated using the mean square error (MSE) given in Equation 15. MSE is generally referred to as precision in the estimation of heterogeneous effect $(PEHE)$, which aside from estimating the CATE, can be used for estimating individual treatment effects as well (ITE) [37-40].

$$\varepsilon_{MAE\_ATE} = |\tau - \hat{\tau}| \tag{14}$$

$$\varepsilon_{PEHE} = \frac{1}{n}\sum_{i=1}^{n}(y_i^1 - y_i^0 - \hat{\tau})^2 \tag{15}$$

The PO Assumptions: The PO framework uses the stable unit treatment value assumption (SUTVA), which deals with the well-defined treatment levels assumption where two instances that are not the same, i.e., $i \not\equiv k$, but have equivalent treatment variable, are thus assumed to receive the same treatment, and the no interference assumption, where a PO of an instance is independent of other instances treatments, written as; $y_j^t = y_j^{t_i}$. Another assumption called the consistency assumption ensures that no multiple versions of the treatment are present in the treatment variable. Thus, it is sometimes referred to as the no multiple treatment assumption [41]. The unconfoundedness assumption is another major assumption employed by the PO framework and it is the focus of our comparative study because it is technically similar to the backdoor adjustment criteria of the SCM.

Definition 8: Unconfoundedness Assumption: The unconfoundedness assumption with respect to a set of covariates in a binary treatment states that the values of the POs are independent of treatment given the set of observed confounding covariates, given as $y_i(1),\ y_i(0) \coprod t_i|x$.

The unconfoundedness assumption is also called conditional ignoreability or just ignoreability, exchangeability, conditional independence [42], or selection on variables [4, 43]. It should be noted that the unconfoundedness assumption is an assumption that cannot be validated unlike the backdoor adjustment criteria of SCM which can be validated with the CIT criteria. Another PO assumption is called the overlap condition (or positivity assumption), which is written as $0 < P(T = 1|x) < 1$. This positivity assumption ensures that dataset strata are fairly distributed along the lines of the treated and the controlled units so that some strata are not populated by only a single treatment unit (treated or controlled). This assumption when combined with the unconfoundedness assumption is called strong ignorability [40, 44, 45], and it ensures that the issue selection bias is properly dealt with, during the analysis and estimation of causal inference in an observation dataset. The assumptions enumerated above are the major assumptions used with the PO framework for identifiability of the required estimand for causal inference in observational studies [12, 36, 46-48]. Thus, in this study, all the assumptions discussed apply implicitly, while we explicitly apply the strong ignorability condition (which combines the unconfoundness and the overlap conditions), for our inference estimand and estimation since all the required covariates in our dataset are measured. The results of the inference estimation under the PO strong ignorability assumptions are presented in section 4.3.

C. Techniques for Adjusting & Estimating Causal Inference:

In this section, we discussed some of the theoretical backgrounds of the techniques for analyzing, adjusting, and estimating the ATE and CATE for both inferences for the SCM and the PO. These estimators are commonly used under the PO framework, and we implemented them empirically for both inferences using our SENSE-EGRA dataset, as shown in section 4. We employed the ordinary least squared (OLS) regression adjustment, propensity score matching (PSM), propensity score weighting (PSW), and propensity score stratification (PSS) under the SCM backdoor adjustment condition and the PO strong ignorability assumption while assuming a binary treatment where $t \in \{0,1\}$. These estimators, which estimate the ATE, the average treatment effects for the controlled (ATC), and the ATT for our dataset adhered to the adjustment criteria identified in our SENSE-EGRA SCM diagram shown in Fig. 4. These estimators as we discussed earlier are capable of dealing with the bias of selection in our dataset, and they are explicated as shown in sections 2.3.1 - 2.3.4.

a) Estimation with Regression Adjustment (OLS):

In supervised machine learning, estimating the probability distribution such as $P(Y, X)$ involves fitting a regression function for the distribution, written as $P(y|x)$, where y is the label and x the features. Albeit in causal effect learning, interventions, and counterfactual (the potential outcome that is not measured) are required, of which two cannot be estimated solely from the dataset. Thus, using the PO framework, we can be able to infer any of the outcomes of the counterfactual C for the treatment t and the covariates set x, as shown in equation 16.

$$C = y_j^{1-t_j} \tag{16}$$

Albeit two types of regression adjustment exist. One involves fitting a single function $P(y|x, t)$ which will suffice for inferring ITE, thus, eliminating confounding bias since the covariates set x is conditioned on. Hence, making the probability distribution equal to the interventional distribution. written as shown in equation 17:

$$P(y|x, t) = P(y|do(T = t), x) \tag{17}$$

Thus by adjusting or conditioning on a unit of treatment $T = t$ and the covariates x as shown in Equation 17, we can guarantee the removal of confounding bias [49] and then further compute the counterfactual as the expected value of the outcome instance y_j conditioned on the control (the converse of the treated, written as $T = 1 - t_j$) and the covariate set x_j of the instance is written as shown in equation 18.

$$C = \hat{y}_j^{1-t_j} = E(y_j|(T = 1 - t_j), x_j) \tag{18}$$

The second regression adjustment involves fitting two models, one representing each PO (i.e., treated and control POs) as shown in Equations 19 and 20.

$$P^1(y|x) = P(y|t = 1, x) \tag{19}$$

$$P^0(y|x) = P(y|t = 0, x) \tag{20}$$

Equations 19 and 20 are regression adjustments for the PO of the treated and control (untreated) respectively. Thus, we can estimate ITE for the PO, written as $\hat{y}_j^t$ using the model $E(y|t, x_j)$ and the ATE $\hat{\tau}$ can further be estimated using Equation 21.

$$\hat{\tau} = \left[\sum_{i=1}^{n} \hat{y}_j^1 - y_j^0\right]/n \tag{21}$$

Where $\sum_{i=1}^{n} \hat{y}_j^1 - y_j^0$ is the summation of the ITEs for the treated minus the controlled (untreated), divided by the sample size n.

b) Adjustment & Estimation with PSS:

The PSS divides variable instances into strata, with each stratum being considered as a simulated RCT, where matches of the treated and controlled are perfectly done, thereby eliminating selection bias. Thus, the naïve estimate of the ATE can be carried out in each stratum [50-52]. In the PSS adjustment estimation, a perfect stratification is assumed, meaning (i) each stratum is defined by a covariate set x; (ii) each stratum covariate set instances x are indistinguishable except for the PO and the treatment [53]. The PSS can be expressed as shown in Equation 22.

$$E\left[y_j^t \middle| t_j = 1, f(x)\right] = E\left[y_j^t \middle| t_j = 0, f(x)\right], \epsilon\ (0,1) \tag{22}$$

where instances are stratified based on the function $f(x)$ and the assumption of strong ignorability holds for each stratum. In specific terms, the ATE for the PSS adjustment is calculated as a weighted average over the group's strata as shown in Equation 23.

$$\hat{\tau} = \sum_i |U_i| \left(\frac{1}{|U_i^1|} \sum_{j \epsilon U_i^1} y_j - \frac{1}{|U_i^0|} \sum_{j \epsilon U_i^0} y_j\right) / \sum_i |U_i| \tag{23}$$

Where the termsU_i,U_i^1 and U_i^0indicates the treatment and controlled instances in the respective strata groups.

However, in the estimation of ATE with the PSS, a condition sometimes occurs when in some strata there only exist instances of the treatment covariates where $t = 1$ or $t = 0$ only and not both. This condition is known as a *lack of overlap*, and when this occurs, the naïve estimator cannot correctly estimate the ATE. To overcome the challenge of *lack of overlap*, the PSW approach is proposed [49], thus, putting aside the need for perfect matching.

c) Estimation using the PSW Adjustment:

PSW is also known as the inverse probability of treatment weighting (IPTW) [54] or double robust estimator (DRE) when regression adjustments are also employed [55]. It weighs treatment instances based on scores obtained from their propensity to mimic an RCT experiment [50]. The weights are expressed as shown in Equation 24.

$$w_j = \frac{t_j}{P(t_j|x_j)} + \frac{1-t_j}{1-P(t_j|x_j)} \tag{24}$$

Thus, for an instance of the treated i and its associate controlled (untreated) instance j, based on Equation 24, the PSW would be $w_i = \frac{1}{P(t_i|x_i)}$ and $w_j =$, which is the inverse probability of getting the observed treated or controlled units conditioned on their covariate set, which is a perfect way to balance the two groups to mimic the RCT or A/B test. Hence, the average weights of factual outcomes for the treated and controlled groups can then be calculated afterward using Equation 25.

$$\hat{\tau} = \frac{1}{n^1}\sum_{j:t_{i=1}} w_i y_i - \frac{1}{n^0}\sum_{j:t_{i=0}} w_i y_i \quad (25)$$

Where n^1, and n^0 indicate the total number of instances under, treated, and controlled (untreated) respectively.

Further, a naïve estimator including a regression adjustment can also be employed in the weighted dataset to minimize the residual of the mimicked A/B test [56]. Albeit, with the PSW method, issues sometimes arise when instances with propensity scores close to 1 or 0 are likely to suffer from having extremely large weights. The solution to this issue is however proposed in the literature by Hernan et al. [57] to mitigate and stabilize the large weights, which we are not able to explicate in this study.

c) Estimation with PSM:

The PSM method is a technique that creates propensity scores based on the weights of covariates in the dataset for both treated and controlled cases. Afterward, it finds matches based on the predicted propensity scores for each similar instance of treated covariates on control covariates to arrive at a balanced set of the treated and controlled covariates so that the ATE can be estimated. The PSM algorithm employs the Greedy one-to-one matching technique for each similar instance of the treated to the controlled [58] units. This ensures that the bias that may be associated with selection during the estimation of the ATE is removed to simulate an A/B test for ATE. The formulae for the estimation of the ATE for the PSM are given in Equation 26.

$$\hat{\tau} = \left[\sum_{i:t_{i=1}}(y_i - y_j) + \sum_{i:t_{i=0}}(y_i - y_j)\right]/n \quad (26)$$

III. EXPERIMENT DESIGN

In this section, we shall discuss the application of the SCM and PO framework in our experiment on the SENSE-EGRA dataset. We explicate the data composition dataset and show the process and procedure (algorithm stages) for the design and validation of the SENSE-EGRA SCM. We also applied the backdoor adjustment condition with the intervention set (the do-calculus) and finally employed the 4 estimators to estimate the causal impact of the SENSE-EGRA intervention program for both the SCM and PO frameworks.

A. SENSE-EGRA SCM Design and the Application of the Backdoor Adjustment Condition:

The SENSE-EGRA dataset on the subtask of letter identification for grade 2 students of two northeast states of Nigeria under study is made of 1,114 records, collected from a population of over 200 primary schools in the said states. 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 instances x stand for $State, LGA, Gender, Age$, etc.; the output feature LI_3 (Y), the treatment variable T $(Treament)$ and two other assessment or evaluation criteria features (LI_1, and LI_2) respectively. See the appendix for more details on the dataset description and encoded meanings.

Thus, based on the above-discussed methodology in section 2 and the conceptual framework guidelines in Fig. 3, we designed the SENSE-EGRA SCM of Fig. 4 using the dataset background knowledge and identified its causal impact based on the four estimators discussed earlier [59, 60]. The following conditional independence criteria are observed and identified in the SENSE-EGRA SCM as shown in Equation 27. The model is validated for correctness with the dataset using the algorithms process described in Fig. 5 and the R package implementation tool of reference [30] and the result is displayed in Table 3.

$$\begin{aligned} &P(LI\_1 \perp X | LI\_2, T \\ &LI\_2 \perp T | X \\ &LI\_3(Y) \perp T | LI\_1, LI\_2, \\ &LI\_3(Y) \perp X | LI\_2, T, \\ &LI\_3(Y) \perp X | LI\_1, LI\_2) \end{aligned} \quad (27)$$

Thus, the estimand for the back-door adjustment criteria, which identified the admissible set of covariates required for adjustment in our SENSE-EGRA SCM of Fig. 4(a) is expressed in Equation 28.

$$P(T, X, LI\_2, LI\_3) = P(LI\_3|X, LI\_2, T) \quad (28)$$

The corresponding NPSEM generated from mutilated DAG (the intervention process) as shown in Fig. 4(b) for our SENS-EGRA SCM designating an intervention distribution is expressed in Equation 29.

$$x = f_x(U_x),\ t = t',\ li_2 = f_{li_2}(x, U_{li_2}), li_1 = f_{li_2}(t, U_{li_1}), li\_3 = f_{li_3}(li_1, li\_2, U_{li\_3}) \quad (29)$$

Notice that in Equation 28, the variable LI_1 is not conditioned on, since from the DAG, it is considered a post-treatment or mediator variable. Pearl et al [1, 3, 4, 31], advised against conditioning on such post-treatment or mediator variables.

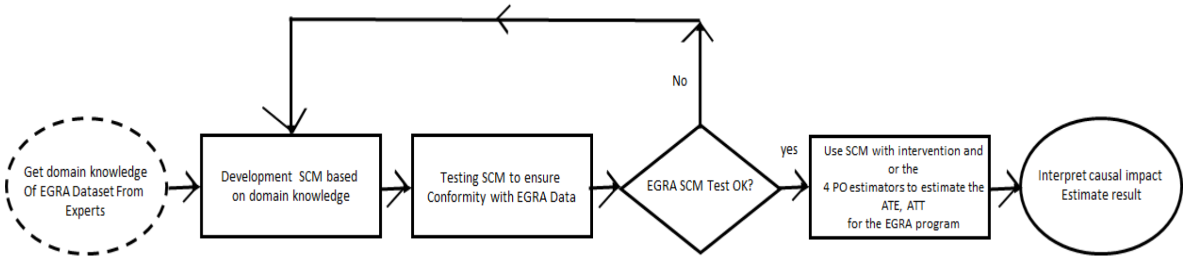
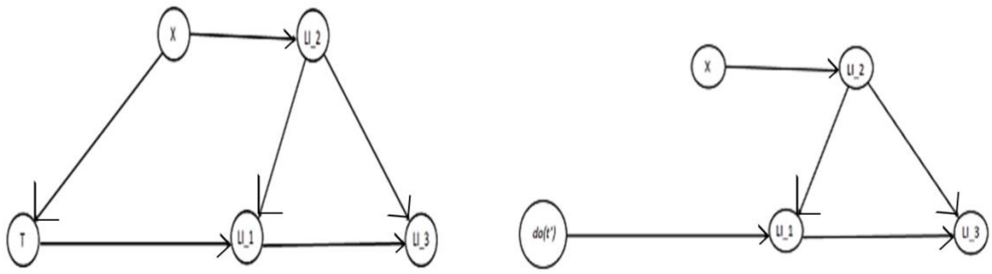


Figure 3: The Conceptual Framework Guideline for Our Mixed Framework for Impact Analysis under Confounding and Selection Biases



(A) An EGRA SCM without intervention (b) An EGRA SCM with an intervention

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

B. Adjustment of the SENSE-EGRA Dataset under the PO Strong Ignorability Assumption:

Under the PO framework, we adopt the unconfoundedness assumption and the positivity or overlap condition, which together are referred to as the strong ignorability assumption for the adjustment of the covariate and the estimation of the causal inference of the SENSE-EGRA intervention program in the area of letter identification. Equation 30 shows the distribution under the strong ignorability assumption.

$$P(T, X, LI\_2, LI\_3) = P(LI\_3|X, LI\_1, LI\_2, T) \quad (30)$$

Notice in Equation 30 that the entire covariates including LI_1, which was not adjusted for under the SCM framework in Equation 28 have been adjusted for, because under the PO framework, there is no DAG to explicitly show the causal relations of variables in the dataset, and no rules exist for how the adjustment should be made except the assumption of the unconfoundedness ab=nd the other proposed which has no validation. Thus, to overcome confounding, all covariates are adjusted for, without recourse to their relationship in the dataset. Section 4 shows the empirical implementation of the results of the PO framework under the strong ignorability and in section 5, a comparative analysis table that shows the difference between the SCM approach and the PO approach is evinced.

IV. RESULTS PRESENTATION & DISCUSSION

In this section, we present the algorithm and the result for the validation of the SENSE-EGRA SCM using the CIT criteria and causal impact estimates under both the SCM and PO frameworks.

A. SENSE-EGRA SCM CIT Validation Results:

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 [1, 2, 18, 61]. One of the most pervasive validation tests for SCM is the use of conditional independence testing (CIT) criteria [1, 61-63]. Surprisingly, in a recent research review paper by Tennant et al., [64], in the testing of DAG of SCM that is implemented in the healthcare sector, for over 200 articles reviewed, not a single one of them tested their SCM for correctness validation. Thus, once the validation process is over, the adjustment criteria can be applied to the SCM. Pearl et al. [1-3], 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 our experiment, we implemented the CIT using the identified conditional independencies set of Equation 27 and applied the back-door adjustment criteria for eliminating confounding bias as shown in Equations 28 and 29 respectively. Table 3 shows the result of the CIT performed on the dataset to verify and validate the correctness of our SENSE-EGRA SCM, implemented in the R package tool of reference [30].

Algorithm 1: Computation of CIT Criteria for any Binary EGRA SCM for Letter Identification Task

1. **Start**
2. **Declare** {X := Set of covariates. Where X ∈ { x_i,..., 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 ∈ {l_i,..., l_n}
   }
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** $\pm$0.1 **then**
   print "CIT validation confirmed"
   else
   print "CIT validation not confirmed"
6. **Repeat** step 3-5:
   for := x_2,..., x_n
7. **End**

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

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}
   }
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** $\pm$ 0.1 **then**
   print "CIT validation confirmed"
   else
   print "CIT validation not confirmed"
6. **Repeat** step 3-5:
   for X := {LGA, School, Gender, Age, Q4, Q5, Q6_0, Q6_1, Q6_2, Q6_3, Q7, Q8, Q9, Q10
   }
7. **End**

Figure 4: Shows the General & And Specific Algorithms Required for the Computations of the CIT criteria for the EGRA and SENSE-EGRA Models for Letter Identification Task

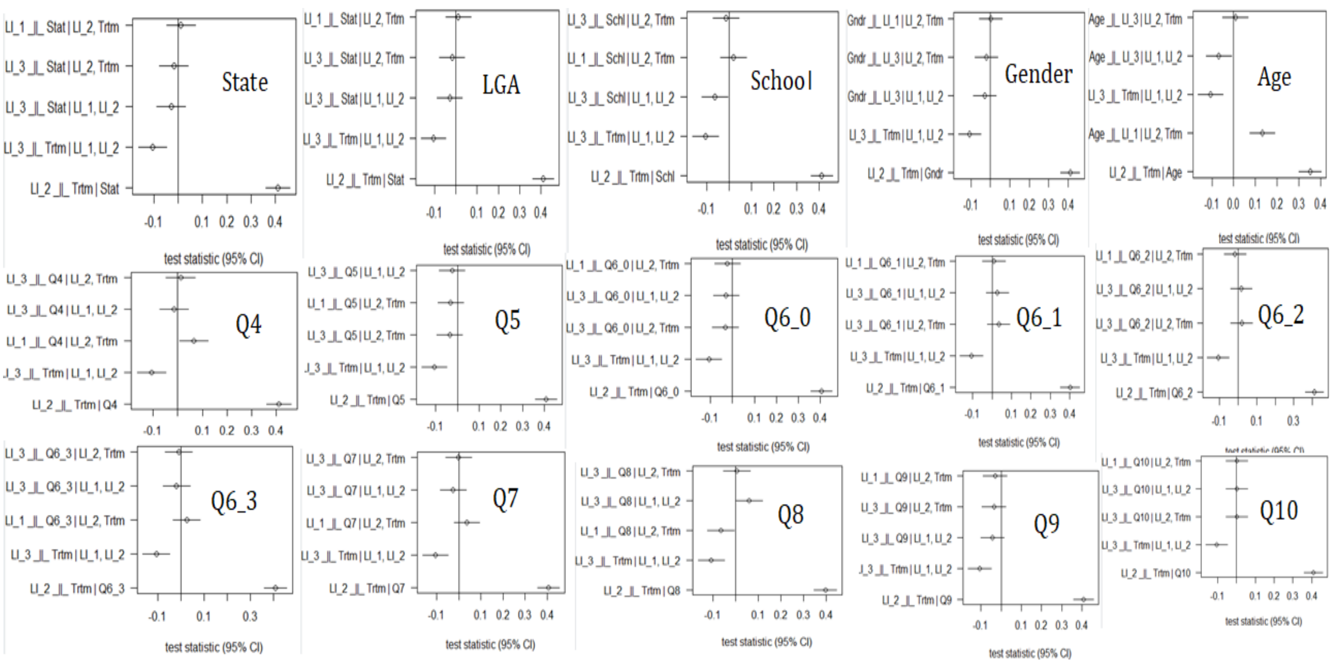


Figure 5: Shows the result of the CIT criteria that plots the local Test results of Table 3 for each variable X

Table 3: shows the result of the CIT identified in Equation 10 for each of the variables X

X =	***Local Tests* for *X***				
	CIT Criteria	**RMSEA**	**p.value**	**2.5%**	**97%**
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	0.05049194	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 ⊥ LI_1|LI_2, T	0.28759887	1.290115e-07	0.07086121	1.1143157
Age	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
	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

a) CIT Condition Results Discussion:

When testing for conditional independence between two or more variables, it is required that their conditional dependency be zero [30]. Hence with the use of the R implementation tool of reference [30], 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. 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 the variables' causal relations [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 Fig. 5.

Thus the R package implementation tool of [30], uses the test functions *LocalTests()* and the plot function *PlotLocalTestResults()* for the computation and analysis of the CIT. The *LocalTests()* function tests the CIT for each feature instance x under the five conditional independence criteria identified in Equation 27, i.e., $X = State, LGA, Gender, Age$, etc., at a confidence interval of 95% for all test cases as shown in column 2 of Table 3. While the *PlotLocalTestResults()* function *plots the results of the localTests() function* as shown in column 3 of Table 3. All the results indicate negative p-values and zero-scale RMSEA values, with the values of each intersection in the plot (*PlotLocalTestResults()*) not exceeding 0.1. Thus, confirming the correctness of our SENSE-EGRA SCM as shown in Table 3 and figure 5.

B. Causal Inference Estimation Results Presentation for SENSE-EGRA program Under the SCM Framework:

Validating the SCM correctness is key in ensuring that the inference that would be estimated from the adjustment criteria of the model would be considered acceptable. Thus, having successfully validated our model's correctness under the CIT criteria. We have also been able to identify the estimand for the backdoor adjustment criteria for the intervention distribution as shown in equations 28 and 29, which effectively handles confounding bias. Thus, to deal effectively with selection bias, we apply the 4 estimation adjustment techniques, which have roots in statistical regression and propensity score, i.e., OLS, PSM (K-Nearest Neighbor(KNNM)), PSW (also called DRE), and PSS (which is blocking) estimators for parametric estimation of the causal inference using the Python Jupyther tools [65-68]. Table 4 shows the result for ATE, the average treatment effect for the controlled (ATC), and the average treatment effect for the treated (ATT), under a 95% confidence interval.

a) Causal Inference Estimates Results Discussion for SENSE-EGRA Program Under the SCM Framework

The causal impact estimates evinced from the OLS and the PSW estimators produced the least and the highest estimates for the ATE as 0.661 and 4.931 respectively, while assuming a constant treatment effect, and thus, making the estimates for ATT, ATC, and the ATE to be the same. The PSS and the PSM estimators, which do not assume constant treatment effects for ATT and ATE, produced more consistent estimates of 2 or approximately 2 for both ATE and ATT as 1.592, 1.535, 2.173, and 1.977 respectively. The seemingly biased estimates evinced by the OLS and the PSW estimators are due to the nonlinearity in the dataset [69, 70] and the presence of the overlap condition [71], which the estimators could not properly handle. Albeit, the PSS and PSM estimators which are not affected by these conditions are biased-free and have more consistent and reliable estimates, which are closed in proximity to each other.

Table 4: Shows the result of the Causal Impact Estimates of the SENSE-EGRA intervention program under the SCM Framework

Estimator int.	Estimate name	Estimate	S.E	Z	95% Conf. Lower	Upper
OLS	ATE	0.661	2.043	0.324	- 3.342	4.665
PSW	ATE	4.931	1.579	3.124	1.837	8.025
PSS (Blocking)	ATE	1.592	1.535	1.037	- 1.416	4.600
	ATC	1.649	2.382	0.692	- 3.020	6.319
	ATT	1.535-	1.721	0.892	- 1.838	4.908
PSM	ATE	2.173	1.279	1.700	- 0.333	4.680
	ATC	2.371	1.467	1.616	- 0.504	5.247
	ATT	1.977	1.645	1.202	- 1.247	5.201

C) Causal Inference Estimation Results Presentation for SENSE-EGRA program Under the PO Framework:

Adjusting and estimating the causal inference for the SENSE-EGRA dataset, using the 4 estimating techniques explained under section 2.3 and the adjustment in equation 30, under the strong ignorability condition of the PO framework, implemented with the Python Jupyther tools [65-68], the following results are elucidated in Table 5.

Table 5: Shows the result of the Causal Impact Estimates for the SENSE-EGRA intervention program under the PO Framework

Estimator	Estimate name	Estimate	S.E	Z	95% Conf. int.	
					Lower	Upper
OLS	ATE	0.079	0.068	1.159	- 0.055	0.213
PSW	ATE	0.035	0.029	1.197	- 0.022	0.092
PSS (Blocking)	ATE	0.036	0.047	0.773	- 0.055	0.127
	ATC	0.013	0.016	0.773	- 0.019	0.044
	ATT	0.060	0.077	0.773	- 0.092	0.211
PSM	ATE	0.066	0.113	0.582	- 0.156	0.288
	ATC	0.036	0.116	0.214	- 0.290	0.361
	ATT	0.096	0.089	1.080	- 0.079	0.272

a) Causal Inference Estimates Results Discussion for SENSE-EGRA Program under the PO Framework:

The causal impact estimates result evinced from the PO framework, showed that the PSW and the PSS produced the most consistent ATE, which are 0.035 and 0.036 respectively, while the OLS and the PSM's ATE also bear close resemblance with each other, i.e., 0.079 and 0.066 respectively, which are the highest ATE estimation results produced. The PSM produced the highest ATT, of 0.096, followed by the PSS with 0.060. The OLS and PSW estimators always assume constant treatment effects. Hence, the ATT and ATE are assumed to be the same, which is 0.079 and 0.035 respectively.

This result shows a great decrease in the results produced by the SCM inference. This is so because the adjustment is made on all the covariates in the dataset including the LI_1 variable which is considered a post-treatment or mediator variable and is considered a forbidden act in the adjustment of covariates in an SCM framework. See Equation 30. Thus, in the SCM where the DAG structure enabled us to identify the variables to perform adjustment on, the PO framework has no such structure or rule preventing covariates adjustment during inference estimation. Also, when covariates are over-adjusted upon the set of covariates, there is a tendency to violate the overlap condition, as some sections of the dataset may contain more or less, or none of their counterpart (treated or controlled units).

V. COMPARATIVE ANALYSIS RESULTS FOR SENSE-EGRA PROGRAM UNDER BOTH SCM AND PO FRAMEWORKS

From the result presentation in Tables 4 and 5, for the SCM and PO framework causal inference result, there is great variation in the result output for the four estimating techniques used, i.e., the OLS, PSW, PSS, and PSM. The SCM causal inference results for the ATE are higher than the PO inference results and are more consistent across almost three of the estimators, i.e., OLS = 0.661, the PSS = 1.592, and the PSM = 2.173, with only the PSW showing an extreme ATE of 4.931. On the other hand, the ATE for the PO causal inference estimates results using the same estimating techniques of OLS, PSW, PSS, and PSM. The result is low when compared to the SCM results, and consistency in ATE result is with only two estimators, i.e., PSW = 0.035 and the PSS = 0.036 respectively, while the OLS and the PSM's ATE also bear close resemblance with each other, i.e., 0.079 and 0.066 respectively. These disparities in the results of the two frameworks are because the adjustment is applied to all the covariates in the dataset including the LI_1 variable in the PO framework, which is considered a post-treatment or mediator variable and is forbidden in the SCM framework for adjustment on [1]. Thus, in the SCM where the DAG structure enabled us to identify the variables to adjust on, the PO framework has no such structure or rule preventing covariates adjustment during inference estimation.

Other differences in the two frameworks outside the causal inferences estimation results advanced above, which differentiates these two frameworks are outlined in Table 6, which is a comparative analysis Table showing the major differences between the SCM and PO frameworks. The summary captured in this table will help new and even old researchers in the field to have good knowledge of them and when to use them in a given situation.

Table 6: Shows Brief Comparative Analysis of the SCM and the PO Frameworks As Used In Our Study

S. No.	SCM Framework	PO Framework
1.	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.
2.	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.
3.	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.
4.	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
5.	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.
6.	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.
7.	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.
8.	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 Table 3 of our study.	The unconfoundedness assumption has no validation.
9.	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.
10.	SCM is used mostly in the field of Computing and related disciplines.	PO is used mostly in social science and econometric disciplines.
11.	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.
12.	SCM was proposed by Judea Pearl a Computer Scientist	PO was proposed by Donald Rubin

VI. CONCLUSION AND FUTURE WORK

This section concludes the study and also takes a peek into our future work.

A. Conclusion:

Causal inference estimates with observational datasets can be a daunting task due to the presence of confounding and selection bias. However, two main frameworks exist for causal inference estimations which are the SCM and the PO frameworks, both of which have techniques or assumptions which they employ in the estimations of causal inference in the presence of confounding bias. The SCM uses a technique called the backdoor adjustment criteria while the PO framework employs the unconfoundedness assumptions and the overlap condition (for dealing with selection bias) which together are called the strong ignorability assumption. Thus, we have used these two methods of the backdoor adjustment and the strong ignorability condition of these two frameworks to empirically carry out causal inference estimation on the SENSE-EGRA observational dataset, an intervention program of the American University of Nigeria, Yola Adamawa state, Nigeria, and using the four estimating techniques of OLS, PSW, PSS, and PSM, which are design to properly handle selection bias for such estimations. Thus, the results show great variation for the SCM and PO frameworks. The ATEs for the SCM are higher than the PO ATE estimates. The ATE for the SCM is more consistent across three of the estimators, i.e., OLS = 0.661, the PSS = 1.592, and the PSM = 2.173, with only the PSW showing an extreme ATE of 4.931. On the other hand, the ATE for the PO evinced is low when compared to the SCM results, and consistency in ATE estimation results is with only two estimators, i.e., PSW = 0.035 and the PSS = 0.036 respectively, while the OLS and the PSM's ATE also bears close resemblance with each other, i.e., 0.079 and 0.066 respectively. These disparities in the results of the two frameworks are due to how the adjustment on covariates is applied. For the PO frameworks adjustment is applied to all the covariates in the dataset including the LI_1 variable, which is considered a post-treatment or mediator variable and forbidden in the SCM framework for adjustment. Thus, in the SCM where the DAG structure enabled us to identify the variables to adjust on, the PO framework has no such structure or rule preventing covariates adjustment during inference estimation. We also presented a table (Table 6) that enumerates other general differences between the two frameworks aside from the difference evinced in our experiment.

B. Future Work:

For future work, we intend to leverage the SENSE-EGRA SCM produced in this study as the basis for establishing causal constraints, from which feasible and actionable local counterfactual explanations of ML predictions, specifically in the task of letter identification predictions of grade 2 student performance can be performed. We believe that these local counterfactual explanations in students' performance predictions of the said EGRA subtask using the causal knowledge gained from the SENSE-EGRA SCM can help build predictive models that would aid stakeholders implementing EGRA intervention programs in the area of letter identification to have a quick and reliable (feasible and actionable) counterfactual explanations of the cause(s) of a student's performance (i.e. good or bad). These local counterfactual explanations will enable the stakeholders responsible for initiating EGRA intervention programs to focus on the key areas for concentration as identified in the model both in the immediate or in the long term for the next intervention program involving the same set of students or similar students elsewhere focusing on the task under review, as the model can easily be adapted and extrapolated to similar datasets.

Availability of Data and Materials:

The dataset used in this experiment can be accessed at the link: https://github.com/Sadaju-Codes/SENSE-EGRA_Project.git.

V. REFERENCES

[1] J. Tian, and J. Pearl, *A general identification condition for causal effects*: eScholarship, University of California, 2002.
[2] J. Pearl, and T. S. Verma, "A theory of inferred causation," *Studies in Logic and the Foundations of Mathematics*, pp. 789-811: Elsevier, 1995.
[3] J. Pearl, and D. Mackenzie, *The book of why: the new science of cause and effect*: Basic books, 2018.
[4] J. Pearl, *Causality*: Cambridge university press, 2009.
[5] C. Cinelli, D. Kumor, B. Chen, J. Pearl, and E. Bareinboim, "Sensitivity analysis of linear structural causal models." pp. 1252-1261.
[6] J. Y. Halpern, "The Book of Why, Judea Pearl, Basic Books (2018)," Elsevier, 2019.
[7] D. B. Rubin, "Causal inference using potential outcomes: Design, modeling, decisions," *Journal of the American Statistical Association,* vol. 100, no. 469, pp. 322-331, 2005.
[8] G. W. Imbens, and D. B. Rubin, "Rubin causal model," *Microeconometrics*, pp. 229-241: Springer, 2010.
[9] D. Aliprantis, "A distinction between causal effects in Structural and Rubin Causal Models," 2015.
[10] D. B. Rubin, E. A. Stuart, and E. L. Zanutto, "A potential outcomes view of value-added assessment in education," *Journal of educational and behavioral statistics,* vol. 29, no. 1, pp. 103-116, 2004.

[11] D. B. Rubin, "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of educational Psychology,* vol. 66, no. 5, pp. 688, 1974.

[12] T. J. VanderWeele, "Ignorability and stability assumptions in neighborhood effects research," *Statistics in medicine,* vol. 27, no. 11, pp. 1934-1943, 2008.

[13] D. F. Heitjan, and D. B. Rubin, "Ignorability and coarse data," *The annals of statistics*, pp. 2244-2253, 1991.

[14] C. A. Flores, and A. Flores-Lagunes, "Identification and estimation of causal mechanisms and net effects of a treatment under unconfoundedness," 2009.

[15] J. M. Wooldridge, "Violating ignorability of treatment by controlling for too many factors," *Econometric Theory,* vol. 21, no. 5, pp. 1026-1028, 2005.

[16] Y. Zhang, A. Bellot, and M. Schaar, "Learning overlapping representations for the estimation of individualized treatment effects." pp. 1005-1014.

[17] V. Karwa, A. B. Slavković, and E. T. Donnell, "Causal inference in transportation safety studies: Comparison of potential outcomes and causal diagrams," 2011.

[18] K. A. Markus, "Causal effects and counterfactual conditionals: contrasting Rubin, Lewis and Pearl," *Economics & Philosophy,* vol. 37, no. 3, pp. 441-461, 2021.

[19] 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.

[20] A. U. o. N. (AUN), "SENSE PROJECT EVALUATION: READING PROFICIENCY ASSESSMENT, RESULTS OF THE HAUSA EARLY GRADE READING ASSESSMENT (EGRA) IN ADAMAWA AND GOMBE," 2022.

[21] S. Bhaskaran, and R. Marappan, "Enhanced personalized recommendation system for machine learning public datasets: generalized modeling, simulation, significant results and analysis," *International Journal of Information Technology,* vol. 15, no. 3, pp. 1583-1595, 2023.

[22] 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," 2020.

[23] S. Taylor, J. Cilliers, C. Prinsloo, B. Fleisch, and V. Reddy, "The Early Grade Reading Study: Impact evaluation after two years of interventions," *EGRS Evaluafion Report*, 2017.

[24] 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.

[25] 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.

[26] B. Piper, and M. Korda, "EGRA Plus: Liberia. Program Evaluation Report," *RTI International*, 2011.

[27] 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.

[28] 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," *International Journal of Educational Development,* vol. 33, no. 3, pp. 283-293, 2013.

[29] 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.

[30] 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.

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

[32] B. Van der Zander, M. Liskiewicz, and J. Textor, "Constructing Separators and Adjustment Sets in Ancestral Graphs." pp. 11-24.

[33] E. Perkovic, J. Textor, M. Kalisch, and M. H. Maathuis, "Complete graphical characterization and construction of adjustment sets in Markov equivalence classes of ancestral graphs," 2018.

[34] P. C. Austin, and E. A. Stuart, "Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies," *Statistics in medicine,* vol. 34, no. 28, pp. 3661-3679, 2015.

[35] P. R. Rosenbaum, and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika,* vol. 70, no. 1, pp. 41-55, 1983.

[36] T. J. VanderWeele, and I. Shpitser, "A new criterion for confounder selection," *Biometrics,* vol. 67, no. 4, pp. 1406-1413, 2011.

[37] J. L. Hill, "Bayesian nonparametric modeling for causal inference," *Journal of Computational and Graphical Statistics,* vol. 20, no. 1, pp. 217-240, 2011.

[38] F. Johansson, U. Shalit, and D. Sontag, "Learning representations for counterfactual inference." pp. 3020-3029.

[39] C. Louizos, U. Shalit, J. M. Mooij, D. Sontag, R. Zemel, and M. Welling, "Causal effect inference with deep latent-variable models," *Advances in neural information processing systems,* vol. 30, 2017.

[40] D. Sontag, U. Shalit, and F. D. Johansson, "Estimating individual treatment effect: Generalization bounds and algorithms," 2017.

[41] S. Weichenthal, S. Ripley, and J. Korsiak, "Fine particulate air pollution and the "no-multiple-versions-of-treatment" assumption: does particle composition matter for causal inference?," *American Journal of Epidemiology,* vol. 192, no. 2, pp. 147-153, 2023.

[42] M. Lechner, "Earnings and employment effects of continuous gff-the-job training in east germany after unification," *Journal of Business & Economic Statistics,* vol. 17, no. 1, pp. 74-90, 1999.

[43] J. Pearl, "Causal inference in statistics: An overview," *Statistics surveys,* vol. 3, pp. 96-146, 2009.

[44] T. Emura, J. Wang, and H. Katsuyama, *Assessing the assumption of strongly ignorable treatment assignment under assumed causal models*, Technical Report of Mathematical Sciences, 2008.
[45] A. Chesher, and E. Battistin, "The impact of measurement error on evaluation methods based on strong ignorability."
[46] R. H. Dehejia, and S. Wahba, "Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs," *Journal of the American statistical Association,* vol. 94, no. 448, pp. 1053-1062, 1999.
[47] R. H. Dehejia, and S. Wahba, "Propensity score-matching methods for nonexperimental causal studies," *Review of Economics and statistics,* vol. 84, no. 1, pp. 151-161, 2002.
[48] T. J. VanderWeele, and O. A. Arah, "Unmeasured confounding for general outcomes, treatments, and confounders: bias formulas for sensitivity analysis," *Epidemiology (Cambridge, Mass.),* vol. 22, no. 1, pp. 42, 2011.
[49] R. Guo, L. Cheng, J. Li, P. R. Hahn, and H. Liu, "A survey of learning causality with data: Problems and methods," *ACM Computing Surveys (CSUR),* vol. 53, no. 4, pp. 1-37, 2020.
[50] P. C. Austin, "An introduction to propensity score methods for reducing the effects of confounding in observational studies," *Multivariate behavioral research,* vol. 46, no. 3, pp. 399-424, 2011.
[51] K. Imai, and M. Ratkovic, "Estimating treatment effect heterogeneity in randomized program evaluation," *The Annals of Applied Statistics,* vol. 7, no. 1, pp. 443-470, 2013.
[52] J. K. Lunceford, and M. Davidian, "Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study," *Statistics in medicine,* vol. 23, no. 19, pp. 2937-2960, 2004.
[53] S. L. Morgan, and C. Winship, *Counterfactuals and causal inference*: Cambridge University Press, 2015.
[54] K. Hirano, G. W. Imbens, and G. Ridder, "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica,* vol. 71, no. 4, pp. 1161-1189, 2003.
[55] M. J. Funk, D. Westreich, C. Wiesen, T. Stürmer, M. A. Brookhart, and M. Davidian, "Doubly robust estimation of causal effects," *American journal of epidemiology,* vol. 173, no. 7, pp. 761-767, 2011.
[56] M. M. Joffe, T. R. Ten Have, H. I. Feldman, and S. E. Kimmel, "Model selection, confounder control, and marginal structural models: review and new applications," *The American Statistician,* vol. 58, no. 4, pp. 272-279, 2004.
[57] M. Á. Hernán, B. Brumback, and J. M. Robins, "Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men," *Epidemiology*, pp. 561-570, 2000.
[58] X. S. Gu, and P. R. Rosenbaum, "Comparison of multivariate matching methods: Structures, distances, and algorithms," *Journal of Computational and Graphical Statistics,* vol. 2, no. 4, pp. 405-420, 1993.
[59] O. Tinubu, A. Sodiya, and O. Ojesanmi, "A behavioral model for characterizing flooding distributed denial of service attacks," *International Journal of Information Technology,* vol. 15, no. 2, pp. 955-964, 2023.
[60] A. Thukral, S. Dhiman, R. Meher, and P. Bedi, "Knowledge graph enrichment from clinical narratives using NLP, NER, and biomedical ontologies for healthcare applications," *International Journal of Information Technology,* vol. 15, no. 1, pp. 53-65, 2023.
[61] J. Pearl, "Models, reasoning and inference," *Cambridge, UK: CambridgeUniversityPress,* vol. 19, no. 2, 2000.
[62] E. Bareinboim, J. Tian, and J. Pearl, "Recovering from Selection Bias in Causal and Statistical Inference (Supplemental Material)," 2014.
[63] E. Bareinboim, and J. Pearl, "Controlling selection bias in causal inference." pp. 100-108.
[64] P. W. Tennant, E. J. Murray, K. F. Arnold, L. Berrie, M. P. Fox, S. C. Gadd, W. J. Harrison, C. Keeble, L. R. Ranker, and J. Textor, "Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations," *International journal of epidemiology,* vol. 50, no. 2, pp. 620-632, 2021.
[65] L. Wong, *Three essays in causal inference*: Stanford University, 2015.
[66] L. Wong, "Causalinference in Python Documentation," 2016.
[67] L. Wong, "Causal Inference in Python," *Causal Inference in Python Blog*, 3rd, Nov, 2022, L. Wong, ed., 2016.
[68] A. Sharma, and E. Kiciman, "DoWhy: An end-to-end library for causal inference," *arXiv preprint arXiv:2011.04216*, 2020.
[69] S. Schwartz, N. M. Gatto, and U. B. Campbell, "Extending the sufficient component cause model to describe the Stable Unit Treatment Value Assumption (SUTVA)," *Epidemiologic Perspectives & Innovations,* vol. 9, no. 1, pp. 1-11, 2012.
[70] A.-G. Dosne, M. Bergstrand, K. Harling, and M. O. Karlsson, "Improving the estimation of parameter uncertainty distributions in nonlinear mixed effects models using sampling importance resampling," *Journal of pharmacokinetics and pharmacodynamics,* vol. 43, pp. 583-596, 2016.
[71] L. A. Aaron, and D. Buchwald, "A review of the evidence for overlap among unexplained clinical conditions," *Annals of internal medicine,* vol. 134, no. 9_Part_2, pp. 868-881, 2001.

Appendix

Appendix: SENSE-EGRA Dataset/Composition/CIT&ATE Codes

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