Combining Dynamic Mode Decomposition and Difference-in-Differences in an Analysis of At-Risk Youth

Marc Andrew Choi, Siyu Huang, Hengyuan Qi, Marco Scialanga
Department of Mathematics
University of California, Los Angeles
inhyak@ucla.edu, nearhuang2019@g.ucla.edu,
qihengyuan@g.ucla.edu, marcoscialanga@g.ucla.edu

Emerson McMullen
Department of Mathematics
Harvey Mudd College
emcmullen@g.hmc.edu

Axel Sanchez Moreno
Department of Mathematics
Texas State University
a_s1350@txstate.edu

Yifei Lou
Department of Mathematical Sciences
University of Texas at Dallas
Yifei.Lou@utdallas.edu

Andrea L. Bertozzi
Department of Mathematics
University of California, Los Angeles
bertozzi@math.ucla.edu

P. Jeffrey Brantingham
Department of Anthropology
University of California, Los Angeles
branting@ucla.edu

2022 IEEE International Conference on Big Data (Big Data) | 978-1-6654-8045-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/BigData55660.2022.10020504

***Abstract*—We analyze the impact of the Los Angeles Mayor's Office of Gang Reduction Youth Development (GRYD) prevention programming using quasi-experimental data. We model the evolution of questionnaire scores and apply Dynamic Mode Decomposition (DMD) to describe the asymptotic behavior of the dynamical system. The analysis indicates that risk decreased for youth who enrolled in GRYD prevention services, while it increased or remained the same for those who were in the control group. We augment these observations using a difference-in-differences (DID) model, showing that the decrease in risk can be attributed to enrolment in prevention services. We draw a connection between DMD and DID using both mathematical analysis and empirical evidence from the questionnaire data. Combining DMD and DID with factor analysis, we investigate the effectiveness of prevention services with respect to different attitudinal domains. We conclude that gang prevention is most effective in impacting attitudes towards negative peer obedience and least effective in impacting attitudes towards violence for self defense. Our analytical approach can be extended to other types of repeated questionnaires.**

***Index Terms*—dynamic mode decomposition, difference in differences, causal inference, factor analysis**

I. Introduction

Gangs remain a difficult challenge for many cities around the world [5]. Starting in 2008, the Los Angeles Mayor's Office of Gang Reduction & Youth Development (GRYD) implemented a comprehensive strategy for community engagement, gang prevention, gang intervention, and violence interruption, all with the goal of helping communities impacted by gang violence [19]. GRYD Prevention, the subject of this study, delivers services to youth between the ages of 10 and 15 who meet certain criteria that place them at-risk of gang involvement. Specifically, youth referred to GRYD by parents, teachers, church figures, or other community members, may complete the Youth Services Eligibility Tool (YSET) questionnaire. The nine-part questionnaire is designed to evaluate attitudes and behaviors related to the risk of joining gangs [6]. If youth exceed certain risk thresholds, they are eligible to receive GRYD Prevention services and take the YSET approximately every six months to assess continued need. Prevention services may include monthly case management team meetings, individual youth and family meetings, intentional youth development activities, and referrals to other services as needed. Prevention services are designed to increase youth and family resilience to the risk factors associated with gang joining [19].

Our goal in this study is to use several mathematical tools to analyze repeated measures of the YSET questionnaire and evaluate impact of GRYD Prevention programming. We start with Dynamic Mode Decomposition (DMD) [15], a data-driven technique, to model the dynamics of time series data. DMD is closely related to the Koopman Operator [12], [16], which approximates a high dimensional nonlinear system into an infinite dimensional linear system. The DMD analysis suggests that GRYD Prevention drives a decline in risk among youth receiving services. By contrast, for a control group of youth who are eligible to receive services, but fail to enroll, DMD identifies an increase in risk over time.

We augment these observations using difference-in-differences (DID) analysis of the same data [22]. DID leverages the control group to model *what would have happened* if the treatment group had not in fact undergone treatment. We obtain statistically significant results from the DID analysis and conclude that the effects of GRYD Prevention are causal. We are able to mathematically connect the results of DMD and DID and establish the conditions under which differences in the leading eigenvalues obtained from DMD are sufficient to reject the null hypothesis that treatment has no effect. This

result suggests that DMD may be further developed for causal analysis.

We also perform factor analysis [17] to understand what latent attitudinal traits affect answers to the questionnaire. We measure the effectiveness of GRYD Prevention on each factor that emerges from the analysis. We observe that GRYD is most effective in mitigating negative peer obedience and least effective in reducing the willingness to use violence for self defense. Using DID, we further investigate the best way to categorize questions. Some but not all questions in factor categories yield negative coefficients in the DID analysis, which suggests the questions designed by GRYD follow a range of different trends under treatment.

The rest of the paper is organized as follows. In Section II, we describe the dataset in more detail. Section III focuses on the theory and results of DMD as it pertains to the YSET data. We conduct causal inference via DID in Section IV, followed by an analysis of the connection between DMD and DID in Section V. In Section VI, we present the results obtained by factor analysis that determine which groups of questions are impacted the most/least by GRYD. We conclude with a summary of the main results.

II. YSET Data

While many of the services provided by GRYD could be beneficial to all youth, the primary goal of GRYD Prevention is to help youth at risk of becoming involved with gangs build resilience. The YSET questionnaire is the primary tool used by GRYD to assess risk and determined eligibility for services. The YSET consists of 104 questions divided into nine different domains corresponding to individual, family and peer risk factors. Risk for any one youth is evaluated relative to threshold measures established in earlier validation studies [6], [7]. Specifically, if the sum of question scores in a target domain exceeds a threshold, the overall risk score is increased by 1. Individuals with a risk score of 4 or higher are eligible for *secondary prevention* (SP) services, the full treatment program. Youth with risk scores between 2 and 3 receive partial services referred to as *primary prevention* (PP). Youth with risk scores below 2 are considered ineligible.

Youth who are eligible and choose to enroll in SP or PP receive services for approximately six months and then retake the YSET. If their risk scores from the retake exceed the established thresholds, youth may continue to receive services, otherwise they "graduate" from GRYD Prevention. YSET instances are labeled as Y_1 for the first intake test for all youth who are referred to GRYD Prevention. Retests for youth who enroll in services are labeled as $R_1, R_2, ..., R_n$, though very few youth remain in services beyond two cycles. Finally, the label Y_2 is used for youth who completed a Y_1 intake test, were eligible for services but, for one reason or another, did not enroll in services. These youth were then referred to GRYD for a second time and completed a new intake YSET. Critically, youth who have only Y_1 and Y_2 results did not receive any Prevention services, though they were eligible after completing the Y_1 intake questionnaire. Assuming that the reasons for not enrolling were random, these youth form a proper control group for comparison with youth who did enroll. To support comparison, we further classify Y_1-Y_2 youth into SP and PP based on their risk scores computed in their first intake test. These are the treatment groups the control youth *would have entered* had they enrolled in services. Table I shows the sample sizes for youth falling into each treatment and control group category.

Table I
Number of Clients in Each Subset of the YSET Data.

	Both PP and SP	SP only	PP only
Y_1-Y_2	70	46	24
Y_1-R_1	6865	3873	2992
R_1-R_2	2149	1737	412
Y_1-R_2	2149	1737	412

A. Data Organization

We focus on six out of the nine domains in the YSET. These categories contain Likert-scale questions relating to antisocial tendencies, weak parental supervision, impulse risk taking, guilt neutralization, negative peer influence, and peer delinquency [7]. There are 31 questions in total, each scored on a 1-5 scale with larger values corresponding to higher risk. Without loss of generality, we normalize each response onto a 0-1 scale. Other domains in the YSET data are excluded from analysis as the questions are either binary true/false, or free response.

As mentioned above, we label the initial intake questionnaire as Y_1, for both treatment and control groups. Every successive retake for the treatment groups are labeled as R_i, $i \in \mathbb{Z}^+$. A second intake questionnaire by eligible youth who do not enroll is labeled as Y_2. We arrange the responses to the intake and each retake questionnaire into a matrix. Each row of a matrix corresponds to a question and each column of the matrix corresponds to an individual's responses to the questionnaire.

In a previous paper [21], the authors applied DMD to an analogous dataset collected for GRYD Intervention Family Case Management (FCM) youth (ages 10-15) and young adults (ages 15-25) who were gang involved. Youth referred to GRYD Prevention may be at risk of gang involvement, but are not gang involved [19]. Several DMD variants, including Exact DMD [20], DMD with control [13], Consistent DMD [2], were investigated in [21], all confirming a decay in risk factors associated with gang involvement. The focus here on youth *at risk* of gang joining, complements this prior work and underscores the comprehensive nature of GRYD programming.

III. DMD Analysis

We aim to analyze the effectiveness of GRYD Prevention services via DMD [20]. We start with the Koopman operator [11], which provides the theoretical basis for DMD analysis. Consider a discrete-time dynamical system

$$x_{k+1} = f(x_k), \tag{1}$$

where x_k, defined on a smooth manifold $\mathcal{M}$, is a state variable at the discrete time k and $f : \mathcal{M} \to \mathcal{M}$ represents the dynamics. In our problem setting, x can be regarded as an individual client of GRYD Prevention. The Koopman operator $\mathcal{K}$ is an infinite-dimensional linear operator that acts on all observable functions $g : \mathcal{M} \to \mathbb{C}$ so that

$$\mathcal{K}g(x) = g \circ f(x). \tag{2}$$

Denote $\varphi_j(x)$ as a Koopman eigenfunction with corresponding eigenvalue λ_j that satisfies

$$\mathcal{K}\varphi_j = \lambda_j \varphi_j. \tag{3}$$

By convention, we assume the magnitude of eigenvalue λ_j is in descending order; i.e., $|\lambda_1| \geq |\lambda_2| \geq \cdots$. A vector of observables (questionnaire scores in our case) $\mathbf{g}$ can be written in terms of Koopman eigenfunctions

$$\mathbf{g}(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_p(x) \end{bmatrix} = \sum_{j=1}^{\infty} \varphi_j(x)\mathbf{v}_j, \tag{4}$$

where $\mathbf{v}_j$ is called the jth Koopman mode. It follows from (2) and (4) that

$$\mathcal{K}\mathbf{g}(x) = \mathcal{K}\sum_{j=1}^{\infty} \varphi_j(x)\mathbf{v}_j = \sum_{j=1}^{\infty} \lambda_j \varphi_j(x)\mathbf{v}_j. \tag{5}$$

As a result, future observables $\mathbf{g}(x_{k+1})$ can be obtained by

$$\mathbf{g}(x_{k+1}) = \mathbf{g} \circ f(x_k) = \mathcal{K}\mathbf{g}(x_k) = \sum_{j=1}^{\infty} \lambda_j \varphi_j(x_k)\mathbf{v}_j. \tag{6}$$

For the response scores at two time points, we can arrange each individual's question scores as a column vector and denote the resulting matrix at $k = 0$ (e.g., intake) by

$$X = [\mathbf{g}(x_0^{(1)}), \cdots, \mathbf{g}(x_0^{(n)})] \in \mathbb{R}^{p \times n}, \tag{7}$$

where $p = 31$ is the number of questions in total to be considered and n is the number of individuals. Similarly we denote the data matrix for the time $k = 1$ (e.g., retake) by

$$Y = [\mathbf{g}(x_1^{(1)}), \cdots, \mathbf{g}(x_1^{(n)})] \in \mathbb{R}^{p \times n}. \tag{8}$$

DMD approximates the Koopman eigenvalues by finding a matrix A such that

$$\mathbf{g}(x_1^{(\ell)}) = A\mathbf{g}(x_0^{(\ell)}), \quad \forall \ell = 1, \cdots, n. \tag{9}$$

Or simply $Y = AX$ using the matrix notation.

The following theorem guarantees that the eigenvalues of A are equivalent to Koopman eigenvalues under certain conditions.

Theorem 1. *Let φ be an eigenfunction of $\mathcal{K}$ with eigenvalue λ. If X defined in (7) is full rank and there exists a vector $\boldsymbol{w} = [w_1, \cdots, w_p]^T$ such that*

$$\varphi(x) = w_1 g_1(x) + \cdots w_p g_p(x) = \boldsymbol{w}^T \boldsymbol{g}(x), \tag{10}$$

then $\boldsymbol{w}^T A = \lambda \boldsymbol{w}^T$, which means λ is a left-eigenvalue of A.

Proof. Applying the Koopman operator $\mathcal{K}$ on both sides of (10) yields

$$\mathcal{K}\varphi(x) = \mathbf{w}^T \mathcal{K}\mathbf{g}(x).$$

Due to $\mathcal{K}\varphi(x) = \lambda\varphi(x) = \lambda \mathbf{w}^T \mathbf{g}(x)$, we obtain

$$\mathbf{w}^T \mathcal{K}\mathbf{g}(x) = \lambda \mathbf{w}^T \mathbf{g}(x). \tag{11}$$

Considering x to be a data point $x_0^{(\ell)} (\ell = 1, \cdots, n)$ in (11), we have

$$\mathbf{w}^T \mathbf{g}(x_1^{(\ell)}) = \mathbf{w}^T \mathcal{K}\mathbf{g}(x_0^{(\ell)}) = \lambda \mathbf{w}^T \mathbf{g}(x_0^{(\ell)}),$$

where the first equality uses the relationship in (5). It further follows from (9) that

$$\mathbf{w}^T A\mathbf{g}(x_0^{(\ell)}) = \lambda \mathbf{w}^T \mathbf{g}(x_0^{(\ell)}).$$

As X is assumed to be full rank, then $\mathbf{w}^T A = \lambda \mathbf{w}^T$. □

Note that Theorem 1 originated from [11, Ch. 3], but with a different assumption that X is full rank rather than $w \in \text{range}(X)$. To make this paper self-contained, we include our proof.

Remark 1. *As indicated in Theorem 1, the Koopman eigenvalue λ is equivalent to the eigenvalue of the DMD matrix A if the set of observables is sufficiently large such that the eigenfunction φ belongs to its span (i.e., $\varphi \in$ span$\{g_1, \cdots, g_p\}$) and the data is sufficiently rich (i.e., X is of full rank).*

Given two data matrices X and Y, the simplest method of finding the matrix A from $Y = AX$ is via the least-squares (LS), i.e.,

$$\hat{A} = \min_A \|Y - AX\|_F^2,$$

which has a closed-form solution,

$$\hat{A} = YX^{\dagger}, \tag{12}$$

where $X^{\dagger}$ denotes pseudo-inverse of X. Due to various nuisance factors (e.g., noise), the LS solution does not give an exact fit, i.e., there exists an error term E such that

$$Y = \hat{A}X + E. \tag{13}$$

To compute DMD eigenvalues (also known as Koopman eigenvalues), the Exact DMD algorithm [20] adopted a rank-reduced singular value decomposition (SVD), which is particularly efficient when the state dimension p is large. Denote the SVD of the matrix $X = U\Sigma V^*$, where $U \in \mathbb{C}^{p \times p}$, $V \in \mathbb{C}^{n \times n}$ are unitary and $\Sigma \in \mathbb{C}^{p \times n}$ is a diagonal matrix. Given a preset value r, reduced SVD approximates the matrix X by $U_r \Sigma_r V_r^*$, where U_r and V_r take the first r columns of U and V, respectively, and Σ_r is an $r \times r$ diagonal matrix. The choice of r

Algorithm 1: Exact Dynamic Mode Decomposition (DMD)

1 Input matrices $X, Y \in \mathbb{R}^{p\times n}$, and an integer $r > 0$
2 Compute the rank-reduced SVD of $X = U_r \Sigma_r V_r^*$
3 Define $\tilde{A} = U_r^* Y V_r \Sigma_r^{-1}$
4 Compute eigendecomposition of $\tilde{A}$, i.e., $\tilde{A}W = W\Lambda$
5 DMD eigenvalues are defined by the diagonal entries of Λ

guarantees that Σ_r is invertible and hence the pseudo-inverse is $X^\dagger = V_r \Sigma_r^{-1} U_r^*$. As summarized in Algorithm 1, Exact DMD computes $\tilde{A} \in \mathbb{R}^{r\times r}$ as a low-dimensional representation of $\hat{A}$ using r most dominant eigenvalues/eigenvectors of $\hat{A}$.

As a data-driven method, DMD computes a finite approximation of the Koopman operator and hence we only have access to a finite number of eigenvalues λ_j in (6) rather than infinite summation. It follows from an induction of (6) that

$$\mathbf{g}(x_k) = \sum_{j=1}^{p} \lambda_j^k \varphi_j(x_0)\mathbf{v}_j, \tag{14}$$

which implies that the leading eigenvalue λ_1 affects the overall trend of the risk scores. In particular, $|\lambda_1| < 1$ results in the decay of $\mathbf{g}(x_k)$ to zero as $k \to \infty$, which can be interpreted as risk reduction. On the contrary $|\lambda_1| > 1$ causes the divergence of $\mathbf{g}(x_k)$, leading to the increase in the risks of joining gangs. The smaller $|\lambda_1|$ is, the greater the decay.

We apply Exact DMD (Algorithm 1) to four pairs of the YSET data, labeled by Y_1-R_1, R_1-R_2, Y_1-R_2, and Y_1-Y_2. We limit ourselves to the setting of $r = p$ since $p = 31$ in our case is computationally tractable. In other words, we use a standard SVD and compute eigendecomposition of $\hat{A}$ completely rather than its lower-dimensional approximation. We visualize the eigenvalues of the four data pairs obtained by Exact DMD in Figure 1, in which x and y axes are real and imaginary components, respectively. We first observe that all the absolute leading (largest in magnitude) eigenvalues have real components less than 1 except for Y_1-Y_2. According to the DMD theory, the questionnaire scores of youth in the SP and PP treatment groups are decreasing in the long run. By contrast, the absolute leading eigenvalue of Y_1-Y_2 is about 1.10, indicating increase in risk as measured by the questionnaire. The comparison of the leading eigenvalues between treatment and control groups implies that GRYD Prevention services are effective at reducing risk. Moreover, the real components of eigenvalues of the Y_1-R_1 group are generally larger than those of R_1-R_2 and Y_1-R_2, suggesting that the GRYD program is more effective over time. We also notice in Figure 1 that the eigenvalues of Y_1-Y_2 are spread out more than those of the other pairs. This may be due to the small data size for Y_1-Y_2.

To gain further insights, we separate the data by youth in SP and PP; i.e., receiving full and partial Prevention services, respectively. The numbers of clients in each group are listed in Table I. The leading eigenvalues of Exact DMD applied on these subsets are recorded in Table II. The leading eigenvalues for clients in SP are generally smaller than the ones for

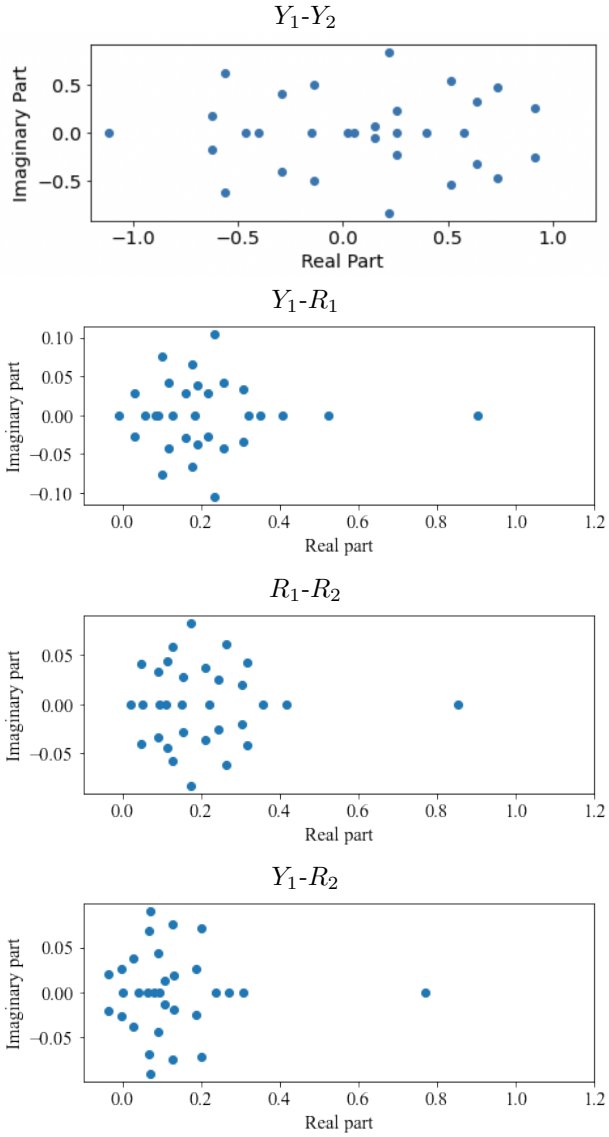


Figure 1. Plots of eigenvalues obtained by Algorithm 1 on both PP and SP data. Note that the leading eigenvalues are smaller than 1, except for Y_1-Y_2 (around 1.10).

Table II
THE ABSOLUTE LEADING EIGENVALUE PRODUCED BY EXACT DMD UNDER DIFFERENT SCENARIOS.

	Both PP and SP	SP only	PP only
Y_1-Y_2	1.10	1.14	NA
Y_1-R_1	0.88	0.84	0.98
R_1-R_2	0.85	0.83	0.96
Y_1-R_2	0.77	0.73	0.92

PP, suggesting that secondary prevention leads to a greater decrease in risk scores than primary prevention.

IV. CAUSAL INFERENCE

In this section, we investigate the causal relationship between enrollment in Prevention services and the decrease in questionnaire scores. The key challenge of any observational

study is that it is impossible to know how a subject *would have responded* had they exposed (or not exposed) to treatment. We only ever observe one or the other potential outcome [23].

Here we apply quasi-experimental difference-in-differences (DID) analysis to compare the outcomes of groups subject to different treatment conditions at different times. DID partitions confounding variables into two categories, namely time-dependent and group-dependent effects. As long as time-dependent effects are the same across groups and group-dependent effects remain fixed over time, it is possible to estimate the effect of treatment [4], [9], [22].

We focus on the simplest DID setting with two groups (control and treatment) and two time points ($T = 0, 1$) that best matches our available data. Specifically, we use the risk scores from (pre-treatment) intake tests ($T = 0$) and (post-treatment) retake tests ($T = 1$) for both those who were enrolled in the GRYD Prevention programming (e.g., Y_1-R_1, SP) and those who were not (e.g., Y_1-Y_2, SP). Define G to be the group indicator, where $G = 0$ refers to the control group of youth who were eligible for services but did not enroll and $G = 1$ for the treatment group who did enroll. Finally, define $D = G \times T$ to be the indicator variable identifying the treatment effect. That is, $D = 1$ only for those individuals who were enrolled in GRYD Prevention Services $G = 1$ during the post-treatment period $T = 1$.

DID can be expressed as linear regression model of *cumulative risk scores*, denoted by Z, with respect to three independent factors T, G, D. The cumulative risk scores are calculated as the sum of each individual question score in the YSET data. Let $\mathbf{t}, \mathbf{g}, \mathbf{d}, \mathbf{z}$ be vectors values of T, G, D, Z for all youth. Mathematically, the DID linear regression model can be expressed as

$$\mathbf{z} = \beta_0 + \beta_1 \mathbf{g} + \beta_2 \mathbf{t} + \beta_3 \mathbf{d} + \boldsymbol{\epsilon}, \tag{15}$$

where $\beta_0, \beta_1, \beta_2, \beta_3$ are scalar coefficients and $\boldsymbol{\epsilon}$ is an error term. Coefficient β_1 accounts for the difference between groups (present when $G = 1$ and absent when $G = 0$); β_2 accounts for differences over time (present when $T = 1$ and absent when $T = 0$). Finally, β_3 accounts for differences attributed to the treatment effect. Specifically, $\beta_3 < 0$ implies that GRYD Prevention services are causally responsible for the decrease in risk. We estimate (15) using linear model estimators in the base R package.

Table I shows the number of individuals falling into treatment (Y_1-R_1) and control (Y_1-Y_2) groups. Table III shows the estimated coefficients for SP and PP separately and combined. The three partitions consistently show that $\hat{\beta}_3$ is negative, and a causal relation is observed for the SP group but not for the PP group. That is, relative to eligible youth who *do not* enroll in Secondary Prevention Services, those who do see significant reductions in risk. Table III also suggests that for both SP and PP youth, group-based and time-based factors exert non-significant influence on risk. Overall, the conclusion is that GRYD Secondary Prevention services lead to significant decreases in risk.

Table III
RESULT OF DID ON CUMULATIVE RISK SCORES

Secondary Prevention	Estimate	p value	t value
$\hat{\beta}_0$: Intercept	14.41	<0.001	25.31
$\hat{\beta}_1$: Group	0.31	0.58	0.55
$\hat{\beta}_2$: Time	1.53	0.06	1.90
$\hat{\beta}_3$: Treatment	-4.98	<0.001	-6.15

Primary Prevention	Estimate	p value	t value
$\hat{\beta}_0$: Intercept	7.34	<0.001	8.82
$\hat{\beta}_1$: Group	0.61	0.46	0.73
$\hat{\beta}_2$: Time	1.53	0.19	1.30
$\hat{\beta}_3$: Treatment	-1.34	0.26	-1.13

Full Data	Estimate	p value	t value
$\hat{\beta}_0$: Intercept	11.99	<0.001	21.12
$\hat{\beta}_1$: Group	-0.21	0.71	-0.37
$\hat{\beta}_2$: Time	1.53	0.06	1.91
$\hat{\beta}_3$: Treatment	-3.40	<0.001	-4.21

In addition to cumulative risk score, we applied DID to each YSET question independently. Here we look at SP data only. The resulting $\hat{\beta}_3$ values are presented in Figure 2. All $\hat{\beta}_3$ values are negative and the estimates for 27 out 31 questions yielded p-values less than 0.05.

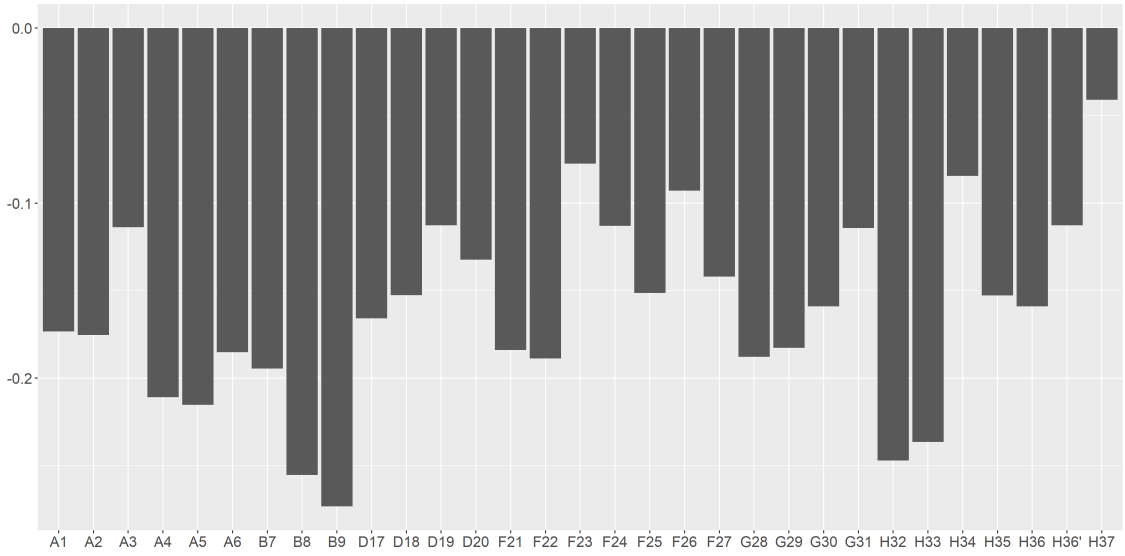
V. CONNECTING DMD AND DID

In Sections III and IV we presented evidence that GRYD Prevention reduces risk for youth receiving services. Whereas DID is a common "work horse" for making causal inferences, DMD typically is not. In this section, we reveal the equivalence between these two models under certain assumptions. Our conclusion is that if the leading eigenvalue of DMD for the control group is larger than the leading eigenvalue for the treatment group by a certain critical value, then the corresponding treatment coefficient $\hat{\beta}_3$ in DID will be negative and statistically significant. We validate this claim empirically on simulation datasets.

We reorganize the response vector $\mathbf{z}$ in (15) as the concatenation of 4 subvectors $\mathbf{z}_{gt}$, each denoting the response variable of group g at time t. Specifically, vectors $\mathbf{z}_{00}, \mathbf{z}_{01}$ are the response variables of the clients in Group 0 (control group) at Times 0 and 1; both are of length n. Similarly, $\mathbf{z}_{10}, \mathbf{z}_{11}$ as response variables of the clients in Group 1 (treatment group) at Times 0 and 1, both are of length m. In other words, $\mathbf{z}$ can be represented as

$$\mathbf{z} = \begin{pmatrix} \mathbf{z_{00}} \\ \mathbf{z_{01}} \\ \mathbf{z_{10}} \\ \mathbf{z_{11}} \end{pmatrix}.$$

Figure 2. Result of DID on Question Scores



Then the DID model (15) can be written as $\mathbf{z} = W\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ and

$$W = \begin{pmatrix} \mathbf{1}_n & \mathbf{0}_n & \mathbf{0}_n & \mathbf{0}_n \\ \mathbf{1}_n & \mathbf{0}_n & \mathbf{1}_n & \mathbf{0}_n \\ \mathbf{1}_m & \mathbf{1}_m & \mathbf{0}_m & \mathbf{0}_m \\ \mathbf{1}_m & \mathbf{1}_m & \mathbf{1}_m & \mathbf{1}_m \end{pmatrix}.$$

The ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (W^T W)^{-1} W^T \mathbf{z}.$$

After calculation, we obtain the closed-form solution

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} \overline{\mathbf{z_{00}}} \\ \overline{\mathbf{z_{10}}} - \overline{\mathbf{z_{00}}} \\ \overline{\mathbf{z_{01}}} - \overline{\mathbf{z_{00}}} \\ (\overline{\mathbf{z_{00}}} - \overline{\mathbf{z_{01}}}) - (\overline{\mathbf{z_{10}}} - \overline{\mathbf{z_{11}}}) \end{pmatrix}, \quad (16)$$

where $\overline{\mathbf{z_{gt}}}$ is average value of the vector $\mathbf{z_{gt}}$.

Now we consider two pairs of matrices: (X_t, Y_t) for the treatment group and (X_c, Y_c) for the control group; each pair containing two data matrices collected at two time points, i.e., (7) and (8).

As $\mathbf{z}_{00}$ is a vector of cumulative risk scores for clients in X_t, we can represent it as $\mathbf{z}_{00} = X_t^T \mathbf{1}_p$, where $\mathbf{1}_p$ denotes the all-one column vector of length p. Then the average of $\mathbf{z}_{00} \in \mathbb{R}^n$, $\overline{\mathbf{z_{00}}}$ can be represented as $\overline{\mathbf{z_{00}}} = \frac{1}{n}\mathbf{1}_p^T X_t \mathbf{1}_n$.

Similarly we have

$$\begin{aligned} \overline{\mathbf{z_{00}}} &= \frac{1}{n}\mathbf{1}_p^T Y_t \mathbf{1}_n, \quad \overline{\mathbf{z_{01}}} = \frac{1}{n}\mathbf{1}_p^T Y_t \mathbf{1}_n, \\ \overline{\mathbf{z_{10}}} &= \frac{1}{m}\mathbf{1}_p^T X_c \mathbf{1}_m, \quad \overline{\mathbf{z_{11}}} = \frac{1}{m}\mathbf{1}_p^T Y_c \mathbf{1}_m. \end{aligned} \quad (17)$$

It follows from the DMD analysis (13) that there exist two matrices A_t, A_c such that $Y_t = A_t X_t + E_t$ and $Y_c = A_c X_c + E_c$, where E_t, E_c are the error terms. We assume A_t, A_c are full rank and hence we can denote the eigenvalue decomposition of A_t by $A_t = \Phi\Lambda\Phi^{-1}$, where $\Phi = [\boldsymbol{\phi}_1, \cdots \boldsymbol{\phi}_p]$ is an invertible matrix composed of normalized eigenvectors and Λ is a diagonal matrix with eigenvalues $\{\lambda_1, \cdots, \lambda_p\}$ on the diagonal. By convention, we assume the eigenvalues are sorted in a descent order in the real components. Similarly we have the eigendecomposition of $A_c = \Psi\Theta\Psi^{-1}$, where $\Psi = [\boldsymbol{\psi}_1 \cdots \boldsymbol{\psi}_p]$ and $\Theta = \text{diag}[\theta_1, \cdots, \theta_p]$.

As both Φ and Ψ are invertible, we define $Z_t := \Phi^{-1} X_t$ and $Z_c := \Psi^{-1} X_c$, which implies that $X_t = \Phi Z_t$ and $X_c = \Psi Z_c$. Combining (16) and (17) with eigendecompositions of A_t, A_c, we can express $\hat{\beta}_3$ by

$$\begin{aligned} \hat{\beta}_3 =& \frac{1}{n}\mathbf{1}_p^T (X_t - Y_t)\mathbf{1}_n - \frac{1}{m}\mathbf{1}_p^T (X_c - Y_c)\mathbf{1}_m \\ =& \frac{1}{n}\mathbf{1}_p^T (\Phi Z_t - \Phi\Lambda Z_t - E_t)\mathbf{1}_n \\ & - \frac{1}{m}\mathbf{1}_p^T (\Psi Z_c - \Psi\Theta Z_c - E_c)\mathbf{1}_m \\ =& \frac{1}{n}(\Phi^T \mathbf{1}_p)^T (I_d - \Lambda)(Z_t \mathbf{1}_n) - \frac{1}{m}(\Psi^T \mathbf{1}_p)^T (I_d - \Theta)(Z_c \mathbf{1}_m) \\ & - \frac{1}{n}\mathbf{1}_p^T E_t \mathbf{1}_n + \frac{1}{m}\mathbf{1}_p^T E_c \mathbf{1}_m, \end{aligned} \quad (18)$$

where I_d denotes the identity matrix. As the matrix $I_d - \Lambda$ is diagonal, we can rewrite

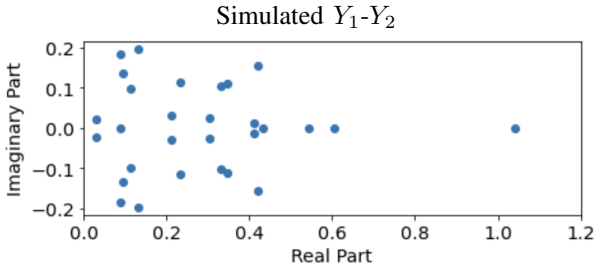


Figure 3. Plots of eigenvalues obtained by Algorithm 1 on simulated Y_1-Y_2 data. The leading eigenvalue is 1.04.

$$
\begin{aligned}
&\frac{1}{n}(\Phi^T\mathbf{1}_p)^T(I_d-\Lambda)(Z_t\mathbf{1}_n)\\
=&\frac{1}{n}\sum_{i=1}^{p}\mathbf{1}_p^T\phi_i(1-\lambda_i)Z_t[i,:]\mathbf{1}_n \ (Z_t[i,:] \text{ is the i-th row of } Z_t)\\
=&\frac{1}{n}\sum_{i=1}^{p}\mathbf{1}_p^T\phi_i(1-\kappa+\kappa-\lambda_i)Z_t[i,:]\mathbf{1}_n\\
=&\frac{1-\kappa}{n}\sum_{i=1}^{p}\mathbf{1}_p^T\phi_i Z_t[i,:]\mathbf{1}_n+\frac{1}{n}\sum_{i=1}^{p}\mathbf{1}_p^T\phi_i(\kappa-\lambda_i)Z_t[i,:]\mathbf{1}_n\\
=&(1-\kappa)\overline{\mathbf{z}_{00}}+\frac{1}{n}\mathbf{1}_p^T\phi_1(\kappa-\lambda_1)Z_t[1,:]\mathbf{1}_n\\
&+\frac{1}{n}\sum_{i=2}^{p}\mathbf{1}_p^T\phi_i(\kappa-\lambda_i)Z_t[i,:]\mathbf{1}_n,
\end{aligned}
$$

which holds for any arbitrary positive real number κ.

Figure 1 and Table II suggest that except for the Y_1-Y_2 group, the leading eigenvalues of both A_t and A_c are real-valued, while their non-leading eigenvalues are clustered and significantly smaller than 1. We can choose a proper κ such that $|\kappa-\lambda_i| \ll 1-\kappa$ for $i=2,\cdots,p$. As a result, we can treat $\frac{1}{n}\sum_{i=2}^{p}\mathbf{1}_p^T\phi_i(\kappa-\lambda_i)Z_t[i,:]\mathbf{1}_n$ as an error term, denoted by e_t. Similarly, we can have $\frac{1}{m}(\Psi^T\mathbf{1}_p)^T(I_d-\Theta)(Z_c\mathbf{1}_m) = (1-\kappa)\overline{\mathbf{z}_{10}}+\frac{1}{m}\mathbf{1}_p^T\psi_1(\kappa-\theta_1)Z_c[1,:]\mathbf{1}_m+e_c$ with an error term e_c. Considering both treatment and control groups, a good choice of κ is the mean of the real components of the non-leading eigenvalues; i.e., $\kappa=\sum_{i>1}(\text{Re}(\lambda_i)+\text{Re}(\theta_i))/(2p-2)$. Ignoring these small error terms (E_t, E_c, e_t, e_c), we get an approximation of (18) as

$$
\begin{aligned}
\hat{\beta}_3 \approx&(1-\kappa)(\overline{\mathbf{z}_{00}}-\overline{\mathbf{z}_{10}})+\frac{1}{n}\mathbf{1}_p^T\phi_1(\kappa-\lambda_1)Z_t[1,:]\mathbf{1}_n\\
&-\frac{1}{m}\mathbf{1}_p^T\psi_1(\kappa-\theta_1)Z_c[1,:]\mathbf{1}_m.
\end{aligned}
\tag{19}
$$

We would like to demonstrate the accuracy of this approximation (19) using simulated Y_1-Y_2 data, as the original Y_1-Y_2 group has a limited number of participants, resulting in more scattered DMD eigenvalues. According to Table III, Y_1-Y_2 shows similar behavior to Y_1-R_1, so we draw $1,000$ samples randomly from the intake responses of the Y_1-R_1 group and take them as the intake responses of the simulated Y_1-Y_2 group. We then predict the retake responses based on the probability distribution of each score of each question in the actual Y_1-Y_2 group. The DMD eigenvalues of the simulated Y_1-Y_2 dataset are presented in Figure 3, which are much more clustered than the ones of the original Y_1-Y_2 shown in Figure 1. The DID result on the simulated dataset is shown in Table IV. Using the full data of Y_1-R_1 and simulated data of Y_1-Y_2, we calculate the right-hand side of (19) to be -4.56, which is a good approximation of the actual value of -4.89, as shown in Table IV. This result suggests that (19) provides a reasonable approximation to the estimated $\hat{\beta}_3$, provided that the DMD algorithm for finding the linear mapping is sufficiently accurate (i.e., $\|E_c\|$ and $\|E_t\|$ are small) and the real components of non-leading eigenvalues are sufficiently close to zero. To the contrary, when these two conditions are not satisfied as in the case using the actual Y_1-Y_2 data, the calculation of $\hat{\beta}_3$ using (19) is 6.03, quite far from the value estimated via DID of -4.89.

Table IV
RESULT OF DID WITH SIMULATED DATA FOR Y_1-Y_2 GROUP

Full Data	Estimate	p value	t value
$\hat{\beta}_0$: Intercept	10.02	<0.001	68.77
$\hat{\beta}_1$: Group	1.75	<0.001	11.24
$\hat{\beta}_2$: Time	3.02	<0.001	14.67
$\hat{\beta}_3$: Treatment	-4.89	<0.001	-22.16

If the treatment and control groups are sufficiently similar in risk scores at pre-treatment (i.e., well-matched), implying that $\overline{\mathbf{z}_{00}} = \overline{\mathbf{z}_{10}}$, then we can compute the following approximations: $\frac{1}{n}\mathbf{1}_p^T\phi_1 Z_t[1,:]\mathbf{1}_n \approx \frac{1}{m}\mathbf{1}_p^T\psi_1 Z_c[1,:]\mathbf{1}_m$, and $\frac{1}{n}\mathbf{1}_p^T E_1\mathbf{1}_n \approx \frac{1}{m}\mathbf{1}_p^T E_2\mathbf{1}_m$. As a result, (19) can be simplified as

$$
\hat{\beta}_3 \approx (\theta_1-\lambda_1)C, \tag{20}
$$

where $C=\frac{1}{n}\mathbf{1}_p^T\phi_1 Z_t[1,:]\mathbf{1}_n=\frac{1}{m}\mathbf{1}_p^T\psi_1 Z_c[1,:]\mathbf{1}_m>0$.

The approximation in (20) reveals a key connection between DMD and DID. That is, the sign of the treatment coefficient in DID is largely affected by the difference of the two leading eigenvalues from DMD for treatment and control groups. If the leading eigenvalue of DMD applied to the treatment group is smaller than the one for the control group, then one can expect $\hat{\beta}_3<0$ in the DID analysis. For example, for the full data, the leading eigenvalue of the simulated control group (1.04 in Figure 3) is larger than the one of the treatment group (0.88 in Table II), so $\theta_1<\lambda_1$, and thus $\hat{\beta}_3<0$. By contrast, applying DID to the full dataset gives $\hat{\beta}_3=-4.89<0$, confirming this result.

This connection between DMD and DID can be made stronger. Specifically given relevant datasets, we can estimate an upper bound of the difference between the two leading eigenvalues, under which the null hypothesis test $H_0:\beta_3=0$ would be rejected. It follows from multiple linear regression theory [1] that the normalized coefficient $\hat{\beta}_3$ follows t-

Table V
CATEGORIES OF THE YSET QUESTIONNAIRE

Label	Category
A	Antisocial Tendencies (A.T.)
B	Weak Parental Supervision (W.P.S.)
DE	Impulsive Risk Taking (I.P.T.)
F	Guilt Neutralization (N.)
G	Negative Peer Influence (N.P.I.)
H	Peer Delinquency (P.D.)

Table VI
THE LEADING EIGENVALUES PRODUCED BY EXACT DMD ON DIFFERENT CATEGORIES.

	A.T.	W.P.S	I.R.T.	N.	N.P.I.	P.D.
Y_1-R_1	0.86	0.80	0.87	0.90	0.84	0.90
R_1-R_2	0.82	0.73	0.82	0.85	0.80	0.82

distribution with degree of freedom $2n + 2m - 4$, i.e,

$$\frac{\hat{\beta}_3}{S_e\sqrt{V_{44}}} \sim t_{2n+2m-4},$$

where S_e is the mean square error (i.e., $S_e^2 = \mathbf{e}^T\mathbf{e}/(2m+2n-4)$ with $\mathbf{e} = \mathbf{z} - \hat{\mathbf{z}} = \mathbf{z} - W\hat{\boldsymbol{\beta}}$) and V_{44} is the (4,4)-entry of the matrix $V = (W^TW)^{-1}$. There is a closed-form formula that gives $V_{44} = \frac{2}{n} + \frac{2}{m}$. For S_e, Equation (16) yields

$$\begin{aligned} S_e^2 = [&(n-1)\operatorname{Var}(\mathbf{z_{00}}) + (n-1)\operatorname{Var}(\mathbf{z_{01}}) + \\ &(m-1)\operatorname{Var}(\mathbf{z_{10}}) + (m-1)\operatorname{Var}(\mathbf{z_{11}})]/ \\ &(2m+2n-4). \end{aligned} \tag{21}$$

By using PP, SP, and full data, we can calculate the variances of $\mathbf{z_{00}}, \mathbf{z_{01}}, \mathbf{z_{10}}$, and $\mathbf{z_{00}}$, leading to an estimate of S_e.

The t-score corresponding to the 0.05 p-value is approximately ± 1.96. As β_3 is negative, it is required that

$$\frac{\hat{\beta}_3}{S_e\sqrt{V_{44}}} < -1.96, \tag{22}$$

to reject the null hypothesis H_0. Combining with (20), we can obtain an upper bound on the difference λ_1 and θ_1 to reject H_0. We still take the full data with simulated control group as an example. We get $C \approx 18, S_e \approx 5, V_{44} \approx 0.0023$. Consequently if $\lambda_1 - \theta_1 > 0.026$, then the hypothesis test H_0 is highly likely to be rejected and a causal relation is established. Note that this threshold value 0.026 is specific to this dataset, as C, S_e, V_{44} all depend on the datasets that we use.

VI. FACTOR ANALYSIS

Finally, we investigate the effectiveness of GRYD Prevention conditioned on the groups of attitudinal traits that the YSET questionnaire seeks to measure.

We start with six categories originally designed by GRYD, labeled as A, B, DE, F, G, and H. As summarized in Table V, each category focuses on a certain risk domain.

To understand which domains GRYD affects the most, we apply Exact DMD on each group of questions and examine the leading eigenvalues. The smaller the leading eigenvalue, the more rapidly the risk scores in that domain decrease.

Table VI presents the leading eigenvalues for two time periods: Y_1-R_1 concerns clients who receive GRYD Prevention services for six months, while R_1-R_2 concerns the same clients who received a second phase of services lasting another six months. The leading eigenvalues show that GRYD is most effective in increasing the importance of parental supervision. Table VI also reveals that GRYD is less effective in lowering risk associated with guilt neutralization[1].

The partition used by GRYD to group questions in the YSET is based on validation research from more than a decade ago [7]. Given the large amount of data now available to GRYD, we can investigate whether alternative groupings of questions are justified. Here we look to data-driven factor analysis (FA) [17] the goal of which is to find a small number of latent factors that account for responses on the observed data.

To determine the optimal number of factors, we use the R package `nFactors` that contains three methods: Kaiser's rule [10], parallel analysis [8] and optimal coordinates [14], returning 9, 8, 8 respectively as the optimal number of factors for the Y_1 data. As Kaiser's rule has a known tendency to over-extract factors [3], we choose 8 as the optimal number of latent variables and employ the R package `factanal` for factor analysis on the Y_1 data. We also allow correlation in these factors by using the optional parameter rotation = "`promax`", as it is reasonable to assume that latent factors in our data are correlated (for example, weak parental supervision should be correlated with peer delinquency). Our factor analysis satisfies all five criteria for a *simple data structure* argued by Thurstone [18] (Figure 4).

Our FA is consistent with the structure used by GRYD for domains B (Weak Parental Supervision), DE (Impulsive Risk Taking), and H (Peer Delinquency). The FA groups questions A1-A5 (Antisocial Tendencies) but places question A6 together with F23, F24, which we identify as attitudes towards "stealing." Factor 7 may be interpreted as "Violence for Defense" due to the prominence of F25 ("It is OK to beat people up if they do something to me first") and F26 ("It is okay to beat people up if I do it to stand up for myself"). Questions grouped in Factor 8 are associated with "Negative Peer Obedience" (Questions G27, G28), while Factor 4 is dominated by Questions G29-G31, regarding "Negative Peer Association." In summary, we note that GRYD's choices in grouping the questions are mostly supported by a data-driven FA approach except for the F (Guilt Neutralization) and G (Negative Peer Influence) categories.

Similarly to Table VI, we apply the Exact DMD algorithm to each group of questions suggested by FA and record the leading eigenvalues in Table VII. Note that the factors S.A., V.D., N.P.O. and N.P.A. are discovered by FA, thus absent in

[1]Two sample questions in the Neutralization category are: "Is it okay for me to lie to someone if it will keep me from getting into trouble with him or her" and "Is it okay to steal something from someone who is rich and can easily replace it."

```
Loadings:
      Factor1 Factor2 Factor3 Factor4 Factor5 Factor6 Factor7 Factor8
A1              0.129                           0.426
A2                     -0.149                   0.487
A3              0.107  -0.123                   0.476
A4                      0.104                   0.351
A5                                              0.657
A6                      0.377                   0.309
B7              0.697
B8              0.868
B9              0.737
DE17                            0.666
DE18                            0.760
DE19                            0.505
DE20                            0.580
F21                     0.170   0.210                   0.161
F22                     0.151   0.165                   0.196
F23                     0.816
F24                     0.862
F25                                                     0.803
F26                                                     0.775
G27                                                             0.769
G28                                                             0.838
G29                                     0.789
G30                                     0.870
G31     0.160           0.145           0.478
H32     0.560                                   0.101
H33     0.532           0.154                   0.114
H34     0.490
H35     0.773
H36     0.784
H36_1   0.529
H37     0.605
```

Figure 4. Loading coefficients for applying factor analysis on the Y_1 data, showing the grouping structure is simple.

Table VII
THE LEADING EIGENVALUE PRODUCED BY EXACT DMD ON EACH FACTOR DETERMINED BY FACTOR ANALYSIS. THE NEW FACTORS (INDICATED BY *) ARE: S.T. : STEALING ATTITUDES; V.D. : VIOLENCE FOR DEFENSE; N.P.O. : NEGATIVE PEER OBEDIENCE; N.P.A.. : NEGATIVE PEER ASSOCIATION;

	A.T.	W.P.S.	I.R.T.	S.A.*	V.D.*	N.P.O.*	N.P.A.*	P.D.
Y_1-R_1	0.86	0.80	0.87	0.81	0.92	0.76	0.83	0.90
R_1-R_2	0.82	0.73	0.82	0.79	0.87	0.77	0.77	0.82

Tables V-VI. Comparing Tables VI and VII, we see that the questions in the original F and G categories do not behave in the same way. Specifically for Category F, GRYD is the least effective on the questions regarding Violence for Self Defense, while the leading eigenvalue associated with Stealing Attitudes is in line with the other factors. We can further differentiate the questions in Category G, as Negative Peer Obedience seems to decrease faster than Negative Peer Association. It seems that GRYD is most effective in decreasing the youths' obedience to friends that affect their lives negatively. On the other hand, GRYD seems to have a weaker impact on the attitudes towards the use of violence for self-defense.

FA suggests that the F and G categories in the YSET could be fine-tuned, since the questions in these original groups do not seem to measure the same latent construct.

VII. CONCLUSIONS

We analyzed temporal changes risk measures for youth referred to GRYD Prevention program, a civilian led approach to helping youth at risk of joining gangs. Using data from the YSET questionnaire administered by GRYD for determining both eligibility and progress in the program, we studied the spectra of the finite dimensional approximations of the Koopman operator obtained by the Exact DMD algorithm. We demonstrated empirically that the leading eigenvalue for two different treatment groups was positive and smaller than 1 and was larger than 1 for the one control group. The results show that questionnaire scores tend to decrease for the youth who are enrolled in GRYD and to increase for those who are not. The data also show a larger decrease for GRYD's Secondary Prevention youth (with full Prevention services) compared with Primary Prevention youth (with partial services), as expected. Furthermore, we performed DID analysis to explore the causal relationship between GRYD Prevention services and the decrease in questionnaire scores. We established empirically a connection between DID and DMD, specifically regarding the leading eigenvalues and the probability to reject the null hypothesis. Finally, we examined a factor analysis of the original questions, suggesting that the F and G domains should be further split into two categories each, as the factors that change the most and the least were hidden with the original partition of the questions given by GRYD.

ACKNOWLEDGMENT

We thank the UCLA CAM REU program for organizing this project. We would also like to thank Andy Chen, Sian Wen, and David Hyde for their guidance. This research was funded by the City of Los Angeles contract number C-132202.

REFERENCES

[1] B. Abraham and J. Ledolter. *Introduction to Regression Modeling*. Duxbury, 2005.

[2] O. Azencot, W. Yin, and A. Bertozzi. Consistent dynamic mode decomposition. *SIAM Journal on Applied Dynamical Systems*, 18(3):1565–1585, 2019.

[3] D. L. Bandalos and M. R. Boehm-Kaufman. Four common misconceptions in exploratory factor analysis. In *Statistical and methodological myths and urban legends*, pages 81–108. Routledge, 2010.

[4] A. Chattopadhyay and J. R. Zubizarreta. On the implied weights of linear regression for causal inference. *arXiv preprint arXiv:2104.06581*, 2021.

[5] S. H. Decker, D. Pyrooz, and J. Densley. *On Gangs*. Temple University Press, Philadelphia, 2022.

[6] K. M. Hennigan, K. A. Kolnick, F. Vindel, and C. L. Maxson. Targeting youth at risk for gang involvement: Validation of a gang risk assessment to support individualized secondary prevention. *Children and youth services review*, 56:86–96, 2015.

[7] K. M. Hennigan, C. L. Maxson, D. C. Sloane, K. A. Kolnick, and F. Vindel. Identifying high-risk youth for secondary gang prevention. *Journal of Crime and Justice*, 37(1):104–128, 2014.

[8] J. L. Horn. A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2):179–185, 1965.

[9] N. Huntington-Klein. *The effect: An introduction to research design and causality*. Chapman and Hall/CRC, 2021.

[10] H. F. Kaiser. The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1):141–151, 1960.

[11] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor. *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.

[12] A. Mauroy, Y. Susuki, and I. Mezić. *Koopman operator in systems and control*. Springer, 2020.

[13] J. L. Proctor, S. L. Brunton, and J. N. Kutz. Dynamic mode decomposition with control. *SIAM Journal on Applied Dynamical Systems*, 15(1):142–161, 2016.

[14] G. Raîche, T. A. Walls, D. Magis, M. Riopel, and J.-G. Blais. Non-graphical solutions for cattell's scree test. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 9(1):23, 2013.

[15] P. J. Schmid. Application of the dynamic mode decomposition to experimental data. *Theoretical and Computational Fluid Dynamics volume*, 11:1123–1130, 2011.
[16] P. J. Schmid. Dynamic mode decomposition and its variants. *Annual Review of Fluid Mechanics*, 54:225–254, 2022.
[17] C. Spearman. General intelligence, objectively determined and measured. *The American Journal of Psychology*, 15(2):201–292, 1904.
[18] L. L. Thurstone. Multiple-factor analysis; a development and expansion of the vectors of mind. 1947.
[19] A. C. Tremblay, D. C. Herz, R. Zachery, and M. Kraus. The city of los angeles mayor's office of gang reduction and youth development (GRYD) comprehensive strategy. *Los Angeles: Mayor's Office of Gang Reduction and Youth Development*, 2020.
[20] J. H. Tu. *Dynamic mode decomposition: Theory and applications*. PhD thesis, Princeton University, 2013.
[21] S. Wen, A. Chen, T. Bhatia, N. Liskij, D. Hyde, A. L. Bertozzi, and P. J. Brantingham. Analyzing effectiveness of gang interventions using koopman operator theory. In *IEEE International Conference on Big Data (Big Data)*, pages 3237–3245. IEEE, 2020.
[22] C. Wing, K. Simon, and R. A. Bello-Gomez. Designing difference in difference studies: best practices for public health policy research. *Annu Rev Public Health*, 39(1):453–469, 2018.
[23] C. Winship and S. L. Morgan. The estimation of causal effects from observational data. *Annual review of sociology*, pages 659–706, 1999.