A Multi-Dimensional Framework for Comparing Digital Learning Effectiveness Between Rural and Urban Higher Education Students

Pranaket Kamble[1], Omkar Nalawade[2], Anurag Mishra[3] and Suraj Talape[4]
Guide: Dr. Prashant Chordiya
DY Patil Institute of MCA and Management, Pune, Maharashtra, India

***Abstract*—**Digital learning is now routine in Indian higher education, yet the evidence on whether rural and urban students benefit equally is mixed. We present DLEF, a measurement-and-decomposition framework with an open reference implementation. It compares rural and urban higher-education students in a single sample on performance impact, engagement, satisfaction, digital literacy, self-efficacy and infrastructure. DLEF scores six instrument dimensions, forms a standardised composite outcome index, tests group differences with Welch and Mann-Whitney tests, Hedges' g, bootstrap intervals and Holm correction, splits the urban-rural gap into explained and unexplained parts with a Blinder-Oaxaca decomposition, and tests a literacy, self-efficacy, engagement pathway by bootstrap mediation. Fifteen automated tests check that the implementation recovers effects planted in synthetic data and reports none where none were planted. Data collection is planned for Pune district, Maharashtra. **This manuscript reports no empirical results about students.** The results section lists the outputs we will report once the collected data are analysed. We draw no conclusion about rural-urban differences here.

***Keywords*—**digital divide, digital learning effectiveness, digital literacy, rural-urban disparity, higher education, Blinder-Oaxaca decomposition, educational technology

I. INTRODUCTION

Digital platforms have moved from a supplement to a main channel of instruction in Indian higher education, a shift accelerated by the COVID-19 lockdowns [1]. National programmes such as Digital India, PM eVIDYA, SWAYAM, DIKSHA and BharatNet were built to make this shift reach every learner. Recent analyses of government and survey data nonetheless report persistent gaps in access, affordability and effective use across rural-urban, gender and income lines, and argue that digital education may reinforce existing inequality by favouring learners who already have devices, connectivity and skills [2]. Whether digital learning narrows or widens the rural-urban gap is therefore still open, and the answer matters for educational equity.

The literature describes the divide at three levels. The first is access. In a survey of 2,592 high-school students in two districts of Karnataka, 20.66% of rural students used computers for academic purposes, against 69.70% of urban students, with power failures and weak computer skills reported as major barriers [3]. The second is skills. A study of 833 university students found that digital competence varied significantly with gender, economic profile and academic level, and that smartphones were the main learning device [4]. The third is outcomes. A survey of 492 Chinese middle-school students found lower behavioural engagement in e-learning among rural students, with e-learning self-efficacy, intrinsic motivation and parental support explaining most of the gap [5]. A study of 245 Pakistani undergraduates likewise linked unequal access and digital literacy to unequal learning outcomes [6]. Together these studies suggest that equal access to devices does not guarantee equal learning.

The evidence for Indian higher-education students is less settled. A large provincial study in China concluded that the device divide had largely closed, with rural students owning phones at higher rates than urban students, and that remaining gaps lay in teacher IT competence, family involvement and independent learning [7]. Studies of Indian undergraduates found no significant difference in digital literacy by locality among 564 students in Coimbatore [8], and none in e-readiness by area among 150 students, 60% of them from rural backgrounds [9]. By contrast, a survey of 1,318 Indian students found significantly higher receptiveness to online learning among urban students [10], and a five-state study of secondary-level English learners reported a large rural-urban performance gap mediated by infrastructure, teacher technology competence and attitudes [11]. These studies differ in population, instrument and outcome, so they cannot be read as a single answer.

From this body of work we identify four gaps. (G1) Studies rarely compare rural and urban students under one design. (G2) Measurement is narrow: access or grades are reported, but performance impact, engagement, satisfaction, literacy and self-efficacy are seldom measured together in one sample. (G3) The contribution of specific factors to the gap is rarely quantified. (G4) Many studies describe the divide without linking it to interventions.

This paper responds to G1-G3 with a measurement and analysis framework, DLEF, and a verified reference implementation, and it states a protocol for the empirical study that will use them. G4 is addressed only in part, as Section III explains. The contributions are:

1) **A framework.** DLEF combines six instrument dimensions, a standardised composite outcome index, group comparison with effect sizes and multiplicity control, an explained/unexplained decomposition of the gap, and a mediation test of one literature-motivated pathway (Section IV).
2) **A study protocol.** Sampling, instruments, hypotheses, power, statistical plan, ablation and sensitivity analyses are specified before any analysis of the collected data (Section V).
3) **A tested reference implementation.** The code is covered by 15 automated tests that check it recovers planted effects in synthetic data and reports none under the null (Section VI).
4) **A synthesis** of 15 studies identifying where evidence agrees and conflicts (Sections I-II).

The paper reports **no empirical findings about students**. Section VIII lists the outputs that will be reported once the analysis is run. The rest of the paper is organised as follows: Section II reviews related work, Section III states the problem and hypotheses, Section IV presents the framework, Section V the protocol, Section VI the implementation, Section VII ethics, Section VIII the results shell, Section IX limitations, and Section X future work and conclusion.

TABLE I
REVIEWED STUDIES (ABSTRACT-LEVEL EXTRACTION)

Ref.	Setting and population	Sample	Finding relevant to this study
[1]	India; students, COVID-19 online education	n.e.	Documents the pandemic-driven shift to online learning
[2]	India; policy and survey data analysis	n.e.	Gaps in access, affordability and use persist; digital education may widen inequality
[3]	Karnataka, India; high-school students, two districts	2,592	Computer use for study 20.66% rural vs 69.70% urban
[4]	Higher education students	833	Digital competence varies by gender, economic profile and level; smartphones main device
[5]	China; middle-school students	492	Rural students lower in behavioural engagement; self-efficacy, motivation, parental support explain most of the gap
[6]	Pakistan; undergraduates	245	Unequal access and digital literacy linked to unequal outcomes
[7]	China; middle-school students, provincial scale	n.e.	Device gap largely closed; gaps remain in teacher IT competence, family involvement, independent learning
[8]	Coimbatore, India; higher education students	564	No significant locality difference in digital literacy
[9]	India; undergraduates (60% rural)	150	No significant area difference in e-readiness
[10]	India; students, pandemic period	1,318	Urban students significantly more receptive to online learning
[11]	Five Indian states; secondary-level English learners	n.e.	Large rural-urban performance gap mediated by infrastructure, teacher technology competence, attitudes
[12]	Online learners in higher education	n.e.	Attitude and literacy predict self-efficacy, which predicts engagement dimensions
[13]	College students	n.e.	Literacy associated with engagement via self-efficacy and learning adaptability
[14]	India; rural-urban education divide (book chapter)	n.e.	Discusses the rural-urban divide in education
[15]	COVID-19 lockdown; multi-layered study	n.e.	Rural-urban divides interact with gender and digital divides

"n.e." = not extracted from the abstract.

II. RELATED WORK

Scope note. This review rests on 15 studies retrieved through a literature-search service, of which only abstracts were read. It is a focused synthesis, not a systematic review, and a full-text pass and a larger corpus are needed before submission.

A. Access and usage. Surveys of Indian students describe large rural-urban differences in whether, and how, computers are used for study [3], and recent policy analyses argue that affordability and connectivity gaps persist across rural-urban, gender and income lines [2]. Student surveys during COVID-19 document the rapid shift to online learning [1]. These studies establish the access level of the divide; they say little about what students learn from the access they have.

B. Skills and readiness. Digital competence is not uniform even among university students, varying with gender, economic profile and level of study [4]. For locality specifically, the Indian undergraduate evidence is mixed: two studies found no significant difference by residence in digital literacy or e-readiness [8], [9], while another found urban students more receptive to online learning [10]. Differences in sample size, instrument and population make the disagreement hard to resolve from the published results alone.

C. Outcomes. Studies that measure outcomes report rural disadvantage in engagement or performance in China [5], Pakistan [6] and India [11], though a provincial study in China found the device gap largely closed and the remaining gaps in teacher competence and independent learning [7]. Most of these studies are not set in Indian higher education, and the Indian one concerns secondary-level English learners [11].

D. Mechanisms. Digital literacy and attitudes predict self-efficacy, which in turn predicts several dimensions of online engagement [12], and self-efficacy and learning adaptability have been reported to mediate the link between digital literacy and engagement [13]. Multi-layered work on South Asia shows that location interacts with gender and social context [14], [15]. These findings motivate measuring several outcomes and testing mediation, rather than comparing grades or access alone.

E. Synthesis. Table I suggests a pattern, which we offer as something to test and not as a conclusion. Studies reporting large rural-urban gaps measured usage or outcomes [3], [5], [6], [11], whereas the three Indian higher-education studies reporting no or mixed locality differences measured literacy, readiness or receptiveness, on samples of 150 to 1,318 [8]-[10]. A difference in what was measured could therefore explain part of the disagreement. That is why we measure access, skills and outcomes together in one sample. The table also shows that the Indian evidence on higher-education outcomes (as opposed to readiness) is thin in this set.

F. What is missing. Across the 15 studies we found none that, within one design, (i) compares rural and urban Indian higher-education students on performance impact, engagement and satisfaction together with literacy, self-efficacy and infrastructure, and (ii) reports how much of the outcome gap is accounted for by differences in those factors. That holds only for the 15 studies we retrieved. Such a study may exist, and Section IX discusses this limit.

III. RESEARCH GAPS, PROBLEM STATEMENT, QUESTIONS AND HYPOTHESES

Gaps and how this paper responds. The four gaps from Section I are addressed to different degrees, and we state this plainly.

TABLE II
RESEARCH GAPS AND THE RESPONSE IN THIS PAPER

Gap	Response in this paper	Extent
G1 Limited comparative frameworks	One design compares rural and urban students on every dimension	Designed, not yet tested on data
G2 Narrow measurement	Six dimensions and a composite index in one sample	Designed; instruments not yet validated
G3 Context-specific factors	Blinder-Oaxaca decomposition of the gap into factor-explained and unexplained parts	Designed, associational only
G4 Little solution orientation	Not addressed; interventions are outside the design	Not addressed

Problem statement. It is unclear whether, and by how much, digital learning effectiveness differs between rural and urban higher-education students in India, and how much of any difference is accounted for by infrastructure, digital literacy and self-efficacy. Published evidence is mixed (Section I), so we treat a rural disadvantage as a hypothesis to be tested, not a premise.

Research questions.

1) RQ1: Does a composite digital learning effectiveness (DLE) index, and each of its components (performance impact, engagement, satisfaction), differ between rural and urban students?
2) RQ2: Do digital literacy, self-efficacy and infrastructure differ between the two groups?
3) RQ3: How much of the DLE gap is associated with differences in those three factors, and how much remains unexplained?
4) RQ4: Is the association between digital literacy and engagement partly carried through self-efficacy?

Objectives. (O1) Specify and verify a reproducible scoring and analysis pipeline. (O2) Validate the instruments in a pilot. (O3) Estimate the group differences, effect sizes and intervals for RQ1-RQ2. (O4) Estimate the decomposition for RQ3 and the indirect effect for RQ4. Only O1 is complete in this paper.

Hypotheses. Because the reviewed studies disagree, tests are two-sided; directions reflect what the stronger studies suggest, not a settled expectation.

TABLE III
HYPOTHESES AND PLANNED TESTS

ID	Hypothesis	Expected direction	Test (Section V)
H1	Mean DLE index differs by group	Urban higher	Welch t, Hedges' g, bootstrap CI
H2	Literacy, self-efficacy, infrastructure differ by group	Urban higher	Same, Holm-adjusted
H3	The explained component of the gap is non-zero	Positive	Oaxaca, bootstrap CI excludes 0
H4	The indirect effect literacy to engagement through self-efficacy is non-zero	Positive	Bootstrap CI excludes 0

The null for H1-H2 (no difference) is a realistic outcome given [8], [9], and a null result would be reported as such.

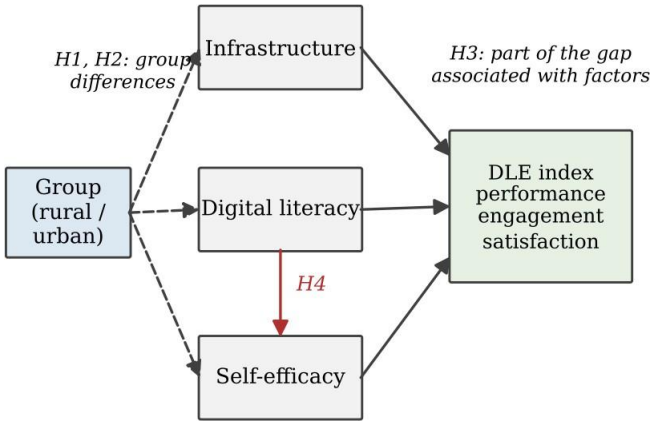


Fig. 1. Conceptual model. Dashed arrows: group differences tested under H1-H2; solid arrows: associations used in the decomposition (H3); red arrow: the mediation pathway (H4). Arrows show hypothesised associations, not established causal effects.

Factors considered and not considered. The framework measures infrastructure, digital literacy and self-efficacy. Two further factors named in the project proposal and in [7], [11], teachers' preparedness for digital instruction and the home learning environment, are only partly covered: home study conditions appear as infrastructure items, but teacher preparedness is not measured. Section IX treats this as a limitation.

Alignment with the Sustainable Development Goals. The study relates to SDG 4 (quality education), because it examines whether digital learning gives comparable benefit across locations, and to SDG 9 (industry, innovation and infrastructure), because infrastructure is one of the factors examined. The study is relevant to both goals, but we do not claim it advances either.

IV. THE DLEF FRAMEWORK

DLEF is a five-stage pipeline from questionnaire responses to hypothesis tests. Every statistical method in it is established; what is proposed is their composition for the rural-urban question, the composite DLE index, and the specific pathway tested. We mark each element as *established* or *proposed*.

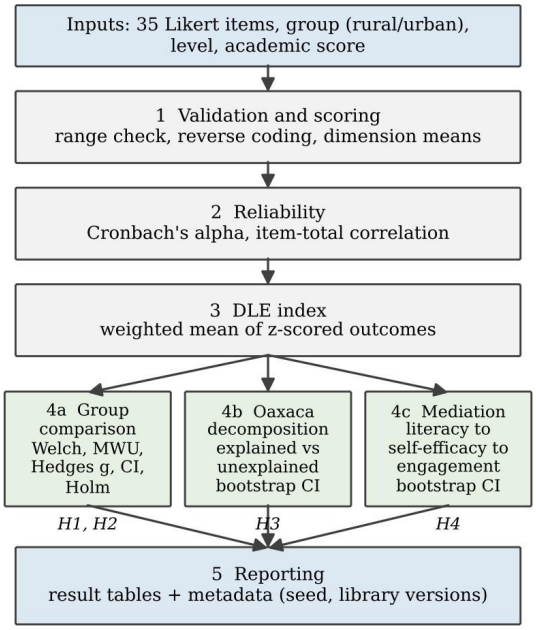


Fig. 2. DLEF architecture and data flow. Boxes 4a-4c run on the output of step 3; each is tied to the hypotheses of Section III.

TABLE IV
NOTATION

Symbol	Meaning
i, n	Respondent index; number of respondents (n_U urban, n_R rural)
k	Dimension index; K_{out} = {performance impact, engagement, satisfaction}
s_{ik}	Score of respondent i on dimension k (mean of recoded items)
z_{ik}, w_k	Standardised score; index weight, non-negative and summing to 1
DLE_i	Composite digital learning effectiveness index of respondent i
X	Factor vector: digital literacy, self-efficacy, infrastructure
β_U, β_R, β^*	OLS coefficients for urban, rural and pooled samples
a, b, c', a·b	Mediation paths and indirect effect
B	Number of bootstrap resamples (2,000 or 5,000)
α, 1 − β	Significance level (0.05) and target power (0.80)

A. Scoring (proposed instrument structure). Six dimensions are scored from 5-point Likert items: performance impact, engagement and satisfaction (outcomes, from the DLEQ), digital literacy and self-efficacy (DLA), and infrastructure (IAI). Item counts are placeholders (5, 6, 5, 8, 5, 6) to be fixed after the pilot. Values outside 1-5 are set to missing, reverse-keyed items are recoded as 6 minus the response, and a dimension score is the item mean. A dimension is left missing for a respondent when more than 20% of its items are missing; no imputation is performed. This threshold is a convention, tested in the sensitivity analysis.

B. Reliability (established). For each dimension with k items, Cronbach's alpha [16] is

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k}\sigma_j^2}{\sigma_T^2}\right) \tag{1}$$

where the σ_j^2 are item variances and σ_T^2 is the variance of the total score. Items with a corrected item-total correlation below 0.30 are flagged, and a dimension is deemed adequate at α of at least 0.70. Both cut-offs are conventional rules of thumb, not guarantees of validity. Alpha does not test dimensionality; Section IX notes the missing factor-analytic step.

C. The DLE index (proposed). Let s_{ik} be respondent i's score on outcome dimension k (performance impact, engagement, satisfaction). With pooled mean and standard deviation, z_{ik} = (s_{ik} minus the mean of s_k) divided by sd(s_k), and

$$DLE_i = \sum_k w_k z_{ik}, \qquad w_k \geq 0, \quad \sum_k w_k = 1 \tag{2}$$

Equal weights are the default. Pooled z-scores put both groups on one scale, so the index is relative to this sample and cannot be compared across studies. The weights are a modelling choice, examined in the ablation (Section V).

D. Group comparison (established methods). For each variable the urban-minus-rural difference is tested with Welch's t-test and the Mann-Whitney U test [17]. Effect size is Hedges' g [18],

$$g = J \cdot \frac{\bar{x}_U - \bar{x}_R}{s_p}, \qquad J = 1 - \frac{3}{4(n_U + n_R) - 9} \tag{3}$$

with pooled standard deviation s_p. A 95% percentile bootstrap interval for the mean difference uses 5,000 resamples [19]. P-values across the seven compared variables (three outcomes, three factors, the DLE index) are adjusted by the Holm step-down procedure [20].

E. Gap decomposition (established method, proposed application). The twofold Blinder-Oaxaca decomposition [21], [22] with the pooled reference coefficients of Neumark [23] splits the mean DLE gap into a part explained by group differences in the factors X (literacy, self-efficacy, infrastructure) and an unexplained part:

$$\begin{aligned} \bar{y}_U - \bar{y}_R &= E + U \\ E &= (\bar{X}_U - \bar{X}_R)^T \hat{\beta}^* \\ U &= \bar{X}_U^T(\hat{\beta}_U - \hat{\beta}^*) + \bar{X}_R^T(\hat{\beta}^* - \hat{\beta}_R) \end{aligned} \tag{4}$$

where E is the explained and U the unexplained component, β* comes from a pooled OLS regression, and the intercept sits in the unexplained part. The explained part is further split by factor. Intervals are 95% percentile bootstrap intervals from 2,000 resamples. The decomposition is associational. It does not show that changing a factor would change the outcome, and the unexplained part absorbs omitted variables as well as any difference in how factors translate into outcomes.

F. Mediation (established method, proposed pathway). With x = digital literacy, m = self-efficacy and y = engagement, the model is m = a·x + e1 and y = c'·x + b·m + e2, and the indirect effect is the product a·b, with a 95% percentile bootstrap interval from 5,000 resamples [24]. The pathway is motivated by [12], [13]. Because the data are cross-sectional, a positive indirect effect would be consistent with the pathway but could not establish the causal order of x, m and y.

G. Power planning (established). With a two-sided test, α = 0.05 and power 0.80, the normal-approximation sample size per group for a standardised difference d is

$$\begin{aligned} n &= 2\left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{d}\right)^2 \\ d_{\min} &= (z_{1-\alpha/2} + z_{1-\beta})\sqrt{2/n} \end{aligned} \tag{5}$$

For d = 0.5 this gives 63 per group. With 150 and 200 students per group the minimum detectable standardised differences are about 0.32 and 0.28. Smaller differences, and most subgroup analyses, will be underpowered.

TABLE V
MINIMUM DETECTABLE STANDARDISED DIFFERENCE (80% POWER, TWO-SIDED ALPHA OF 0.05)

Students per group	Total	Minimum detectable standardised difference
50	100	0.56
75	150	0.46
100	200	0.40
150	300	0.32
200	400	0.28

The values come from the formula above (80% power, two-sided alpha of 0.05, normal approximation) and assume independent respondents. Students clustered in institutes carry less information than independent ones: the required sample is inflated by the design effect 1 + (m - 1)ρ, where m is the average cluster size and ρ the intraclass correlation. ρ is unknown and is to be estimated from the pilot, so the figures above are lower bounds on the detectable difference.

H. Algorithm and complexity. Algorithm 1 summarises the pipeline as implemented.

Algorithm 1: DLEF analysis pipeline

```
Input: survey table D (n rows), weights w, seed, B
Output: reliability, comparison, decomposition, mediation
    tables
1:  D <- set Likert values outside 1..5 to missing
2:  for each dimension d: recode reverse-keyed items (6 -
    x)
3:  s_d <- mean of items; set missing if more than 20%
    items missing
4:  for each outcome dimension k: z_k <- (s_k - mean) /
    sd (pooled)
5:  DLE <- sum_k w_k * z_k
6:  reliability <- alpha_d and item-total correlations
    for all d
7:  for each variable v in {outcomes, factors, DLE}:
8:    test urban vs rural (Welch, Mann-Whitney); g_v;
      bootstrap CI (B)
9:  p_adj <- Holm(p over all v)
10: (gap, explained, unexplained) <- Oaxaca(DLE ~
    factors), bootstrap CI (B)
11: (a, b, c', ab) <- mediation(literacy, self-efficacy,
    engagement), bootstrap CI (B)
12: write tables and metadata (seed, versions)
```

Scoring and reliability are linear in the number of items times respondents, O(n·m). Each bootstrap costs one resampling plus an OLS fit, so decomposition and mediation scale as O(B·n·p^2) for p regressors (p = 4 including the intercept), which is small for the sample sizes planned. The runtime is dominated by the resampling loops; measured runtime and memory are to be reported with the hardware used (Section V-I).

I. Components and data flow. Five components are chained, each reading the output of the one before. (1) *Intake and validation* receives the response table, checks that every expected column exists and sets out-of-range Likert values to missing. (2) *Scoring* turns items into six dimension scores and the DLE index and passes on a score table. (3) *Reliability* reads the item-level data and writes alpha and flagged items per dimension; it does not change the data. (4) *Inference* reads the score table and the group label and writes the comparison, decomposition and mediation tables. (5) *Reporting* writes the tables and a metadata file with the seed and library versions. Raw responses are never modified, every step is deterministic given the seed, and each written table can be re-read and checked independently.

J. Design rationale. A single composite index lets the primary hypothesis be tested once, without a multiplicity penalty, while the component analyses stay available as secondary tests. The Blinder-Oaxaca decomposition was chosen because it answers the question the literature leaves open, namely how much of a group gap is associated with measured factors, and because it needs only regression. Mediation is restricted to one pathway fixed in advance from [12], [13], to limit forking paths. DLEF deliberately contains no machine-learning model. With a sample of a few hundred, a statistical design keeps every estimate interpretable and reproducible; predictive models, for example regularised regression for factor importance, would need larger samples and are left as an extension.

TABLE VI
COMPARISON OF APPROACHES IN THE REVIEWED LITERATURE AND DLEF

Approach	Measures	Explains the locality gap?	Examples
Access or usage surveys	Devices, connectivity, computer use	No	[3]
Readiness or literacy comparisons	Digital literacy, e-readiness, receptiveness	No	[8]-[10]
Outcome comparisons with regression or mediation	Engagement or performance plus mediators	Partly, through mediators	[5], [11]
Mechanism studies	Attitude, literacy, self-efficacy, engagement	Pathway, not the locality gap	[12], [13]
DLEF (proposed)	All six dimensions in one sample	Decomposes it (associational)	This paper

V. STUDY DESIGN AND EVALUATION PROTOCOL

A. Design and population. The study is quantitative, non-experimental and cross-sectional. The target population is undergraduate and postgraduate students in Pune district, Maharashtra, India, with at least six months of digital-learning exposure. *Urban* means residence within the Pune or Pimpri-Chinchwad municipal areas or the cantonment; *rural* means residence in a Pune-district taluka outside them. Group is assigned from verified residence at intake, not from the location of the institute.

B. Sampling. The protocol specifies stratified random sampling of 300-400 students, split evenly between groups and stratified by level (UG/PG), drawn from a sampling frame of participating institutes (the implementation allocates quotas within each group in proportion to stratum size and fails rather than under-sample). Respondents are recruited by stratified random sampling from this frame. If recruitment departs from this design, for example through the authors' own contacts, the sample is a convenience or snowball sample. Population inference to Pune district or India would then be unsupported, the analysis must be reported as exploratory, and the abstract and conclusions must say so.

C. Instruments. The DLEQ (performance impact, engagement, satisfaction), DLA (literacy, self-efficacy) and IAI (infrastructure) are newly written for this study; their items are not yet validated. Academic records (most recent term percentage) and short semi-structured interviews are used for validation. A collection template with data validation (Likert range, group, level, taluka list) is provided with the code.

TABLE VII
DRAFT INSTRUMENT ITEMS (PROPOSED, UNVALIDATED; FOR THE PILOT)

Dimension (instrument)	Draft items
Performance impact (DLEQ)	(1) Digital learning helps me understand course content better. (2) My academic performance has improved since digital learning began. (3) I complete coursework more efficiently using digital resources. (4) Digital learning helps me retain what I study. (5) I can apply what I learn digitally to practical tasks.
Engagement (DLEQ)	(1) I take part regularly in online classes and discussions. (2) I attend digital sessions attentively. (3) I complete online activities on time. (4) I ask questions or seek help during digital learning. (5) I stay motivated to study on digital platforms. (6) I explore digital resources beyond what is required.
Satisfaction (DLEQ)	(1) I am satisfied with the quality of digital instruction. (2) The platforms used are easy to use. (3) I am satisfied with the feedback I receive digitally. (4) Digital learning suits my study routine. (5) I would like digital components of my course to continue.
Digital literacy (DLA)	(1) I can find reliable academic information online. (2) I can judge whether an online source is trustworthy. (3) I can use word-processing, spreadsheet or presentation tools for coursework. (4) I can join and manage online classes or meetings. (5) I can submit assignments through an online portal. (6) I can protect my accounts and personal data. (7) I can fix common problems with my device or apps. (8) I can acknowledge and cite online sources.
Self-efficacy (DLA)	(1) I am confident learning new digital tools. (2) I can complete online tasks without help. (3) I believe I can succeed in digital courses. (4) I can handle technical difficulties on my own. (5) I am confident using digital tools for assessments.
Infrastructure (IAI)	(1) I have reliable access to a suitable device for study. (2) My internet connection is stable enough for live classes. (3) My electricity supply is rarely interrupted while I study. (4) I have a quiet place at home to study digitally. (5) Internet cost does not limit my study. (6) My institute provides adequate digital facilities.

All items are 5-point agreement scales (1 = strongly disagree, 5 = strongly agree). No item is reverse-keyed in this draft. Dimension scores are item means.

The literacy dimension as drafted is a self-rating of skills, not a performance test, which limits what it can show (Section IX). Items will be reviewed by at least two subject experts before the pilot.

D. Pilot (Phase 2). At least 30 respondents per group complete the draft instruments. Outputs: alpha and flagged items per dimension, completion time, and comprehension notes. Items below the thresholds in Section IV-B are revised or dropped, and the pilot respondents are excluded from the main sample.

E. Main analysis. The primary analysis is H1 (DLE index difference) at $\alpha = 0.05$. H2 comparisons are Holm-adjusted across the seven variables. H3 and H4 are reported with bootstrap intervals. Missing data are handled by the 20% rule in Section IV-A and complete-case analysis otherwise, and the number of rows lost is reported. All analysis choices above are fixed before the main data are examined.

F. Baseline analyses. No prior method estimates the same quantities, so there is no published model to benchmark against. Three simpler analyses serve as comparison for what the multi-dimensional design adds: (B1) a grades-only comparison using the academic score, (B2) an access-only comparison using the infrastructure score, and (B3) the group comparison without decomposition. Comparison with published studies is qualitative, because populations and instruments differ.

G. Validation checks. Convergent validity: Spearman correlation between the performance-impact score and the academic score. Qualitative validation: whether interview themes agree with the survey direction for the same respondents.

H. Ablation and sensitivity plan.

TABLE VIII
ABLATION AND SENSITIVITY PLAN

Analysis	What is varied	Question answered
Index weights	Equal weights vs each outcome weighted 0.5	Does H1 depend on the weights?
Component removal	Drop one outcome dimension from the index	Is any one dimension driving the gap?
Factor set in decomposition	Drop each factor in turn	How stable is the explained share?
Missing-data rule	Threshold 0%, 20%, 40%	Do conclusions depend on the 20% rule?
Test choice	Welch vs Mann-Whitney vs bootstrap	Do conclusions depend on distributional assumptions?
Sample size and noise	Subsampling; simulated missingness	How fast does power degrade? (script provided)

I. Runtime and resources. The pipeline is lightweight. A benchmark script measures runtime and peak memory against sample size, and will report them with the hardware used. It has not been run on study data.

J. Planned schedule. Weeks 1-4 literature review and instrument development; weeks 5-6 pilot and validation; weeks 7-10 data collection; weeks 11-16 analysis and report.

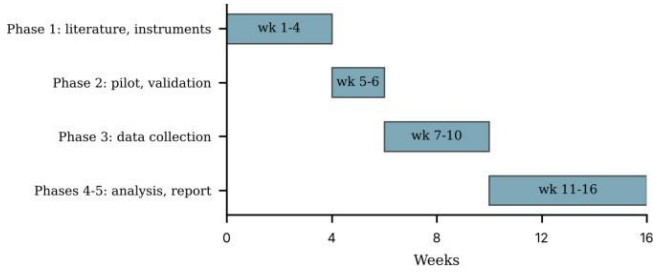


Fig. 3. Planned 16-week schedule.

K. Variables.

TABLE IX
VARIABLES AND OPERATIONAL DEFINITIONS

Variable	Type	Operational definition	Source
Group	Binary	Rural or urban, from verified residence	Intake
Level	Binary	Undergraduate or postgraduate	Intake
Months of digital learning	Integer	Eligibility requires at least 6	Intake
Performance impact	1-5	Mean of 5 DLEQ items	DLEQ
Engagement	1-5	Mean of 6 DLEQ items	DLEQ
Satisfaction	1-5	Mean of 5 DLEQ items	DLEQ
Digital literacy	1-5	Mean of 8 DLA items (self-rated)	DLA
Self-efficacy	1-5	Mean of 5 DLA items	DLA
Infrastructure	1-5	Mean of 6 IAI items	IAI

Variable	Type	Operational definition	Source
DLE index	z-units	Weighted mean of the three z-scored outcomes	Derived
Academic score	0-100	Latest-term percentage from institute record	Records

L. Decision rules and error analysis. A result is called statistically significant at the Holm-adjusted 0.05 level for secondary comparisons and the unadjusted 0.05 level for the single primary test (H1). The Holm family is the seven variables of Section IV-D, so the DLE index also appears among the adjusted values; its unadjusted p-value is the one used for H1. Practical importance is judged from Hedges' g and the confidence interval, not from p alone, and a difference whose interval spans values both trivial and large is described as imprecise. For error analysis, respondents are checked for straight-lining (identical answers across all items), implausible completion times, and extreme index values. Analyses are repeated without flagged respondents and any change is reported. Residuals of the decomposition regressions are inspected for non-linearity and heteroscedasticity.

M. Reproducibility checklist.

TABLE X
REPRODUCIBILITY CHECKLIST

Item	Plan	Status
Code	Package, runner and tests with pinned dependencies	Done
Seeds and versions	Fixed seeds; versions written to run metadata	Done
Analysis plan	Hypotheses, tests and thresholds fixed before main-data analysis	This paper
Data dictionary	Column definitions, allowed values, units	Done
Collection template	Validated Excel template with dropdowns and Likert limits	Done
Raw data	Not shared (participant privacy); aggregate tables shared	Pending ethics
Hardware and runtime	To be reported with benchmark output	Pending
Deviations	Any change from this protocol reported with reason	Ongoing
Repository	Public deposit with persistent identifier on acceptance	Pending

N. Administration and quality control. The instruments are prepared in English and in Marathi, translated forward and back by two bilingual reviewers, and tried on 5-8 students in cognitive interviews to check that items are understood as intended. The survey is offered on paper as well as online. This matters for validity: an online-only survey would systematically miss rural students with poor connectivity, the very respondents whose experience the study is about. Each respondent completes one form, identified only by a random code, and duplicate submissions are checked by code and institute. Data entry from paper forms is done twice and compared. Interviews (about 10 per group, continued until no new themes appear) are semi-structured, audio-recorded only with consent, and transcribed without names.

VI. REFERENCE IMPLEMENTATION AND VERIFICATION

The implementation is a small Python package, with dependencies pinned in a requirements file (numpy 2.4.4, scipy 1.17.1, pandas 3.0.2, pytest 9.1.1). All resampling uses fixed seeds, and the runner writes the seed and library versions beside its output tables.

TABLE XI
IMPLEMENTATION MODULES AND STATUS

Module	Function	Status
config	Dimension and instrument schema	Implemented
sampling	Stratified random sampling, equal allocation by group	Implemented, tested
scoring	Range check, reverse coding, dimension scores, DLE index	Implemented, tested
reliability	Cronbach's alpha, item-total correlations, flags	Implemented, tested
stats	Welch, Mann-Whitney, Hedges' g, bootstrap CI, Holm, power	Implemented, tested
decomposition	Blinder-Oaxaca with bootstrap CI	Implemented, tested
mediation	Product-of-coefficients with bootstrap CI	Implemented, tested
run_study	End-to-end runner writing result tables	Implemented, smoke-tested
robustness, benchmark	Sensitivity and runtime scripts	Written, not yet executed
Spearman validation, factor analysis, multilevel models	Planned in Sections V and IX	Not implemented

Verification. Fifteen automated tests pass. They check known-answer cases (reverse coding, alpha for perfectly correlated and for random items, Holm adjustment, Hedges' g of an identical sample, the power formulas), error handling (out-of-range values, over-missing items, a sampling quota that cannot be met, invalid weights), the algebraic identity that the explained and unexplained parts sum to the gap, and recovery tests on synthetic data generated with arbitrary parameters: the pipeline detects a group gap and an indirect effect that were planted, attributes a planted direct gap to the unexplained part, and, when none was planted, returns a group-difference p-value above 0.01 and an indirect-effect interval that contains zero. Each recovery test uses one fixed seed, so these are existence checks, not estimates of error rates. One defect in the synthetic generator was found and fixed during testing.

What this does and does not show. The tests show that the code computes what Section IV specifies, and that it behaves sensibly where the truth is known by construction. They do not show that the instruments measure what they claim to, that the DLE index is a valid measure of learning effectiveness, or anything about real students. The synthetic parameters are test fixtures and are not reported as findings.

VII. ETHICAL CONSIDERATIONS AND DATA SECURITY

The study involves human participants and needs institutional ethics clearance before data are analysed or published. Ethics clearance has not yet been applied for. It will be sought from

the institute's ethics committee, and no participant data will be collected until it is granted.

- **Consent.** Participation is voluntary and based on informed consent recorded before the survey. Participants may skip any item or withdraw before analysis. Only adults (18 or over) are eligible.
- **Recruitment.** Where respondents come through the authors' friends, colleagues or their contacts, there is a risk that people feel obliged to take part. The invitation must say that refusal has no consequence, and no respondent's answers may be shown to anyone who knows them.
- **Data minimisation.** No names, phone numbers, email addresses or exact addresses are collected. Age is held in bands, location as institute code and area or taluka, and each respondent has a random identifier. Linkage between identifiers and any contact record, if one exists for follow-up interviews, is kept separately and destroyed after interviews.
- **Security.** The raw dataset will be held in access-restricted storage, shared only among the named authors, never uploaded to third-party services without consent, and reported only in aggregate. Cells with very few respondents are not reported.
- **Synthetic data.** The synthetic test data used to verify the code is generated by program, labelled as such in every row, and is kept separate from study data.
- **Interviews.** Interview notes are coded to themes and quoted only after removing details that could identify a person or institute.

The study does not include animal or clinical components, and no intervention is applied to participants.

VIII. RESULTS

No empirical results are reported in this version. The pilot and main surveys have not been analysed in this manuscript. The table below lists what the pipeline writes and which hypothesis each output answers, so the section can be completed directly from the output files without changing the analysis.

TABLE XII
PLANNED OUTPUTS AND THE HYPOTHESES THEY ANSWER

Output file	Content	Answers	Status
reliability	Alpha and flagged items per dimension (pilot, then main)	Instrument adequacy	Pending data
group_comparison	Means, difference, 95% CI, Hedges' g, raw and Holm p for seven variables	H1, H2, baselines B1-B2	Pending data
oaxaca	Gap, explained, unexplained, explained share per factor, 95% CI	H3, baseline B3	Pending data
mediation	Paths a, b, direct c', indirect a·b, 95% CI	H4	Pending data
validation	Spearman correlation with academic score; interview agreement	Validity	Not implemented
robustness, benchmark	Detection rates under noise and missingness; runtime and memory	Section V-H, V-I	Written, not run
run_metadata	Seed, library versions, rows read and retained	Reproducibility	Written by the runner

Reporting rules once results exist. Every number is copied from the output files. Null and negative results are reported. Directions are reported against the two-sided tests of Section III. Exploratory analyses are labelled as such. If the sample is a convenience sample (Section V-B), results are described as applying to the respondents and not to rural or urban students in general.

TABLE XIII
INTERPRETATION GUIDE (FIXED IN ADVANCE; NO RESULT IS IMPLIED)

If the output shows	It may be read as	Caution
Interval for the DLE difference excludes 0; g of about 0.2 or more	Evidence of a group difference in this sample	Direction and size must be reported; trivial g is not a meaningful gap
Interval includes 0 and is narrow	Evidence of little or no difference of practical size	Consistent with [8], [9]; report the interval
Interval includes 0 and is wide	Inconclusive; the sample cannot say	Underpowered for small effects (Section IV-G)
Welch and Mann-Whitney disagree	Sensitivity to distribution shape	Report both; prefer the more conservative
Large explained share in the decomposition	The gap is statistically associated with factor differences	Not causal; does not show an intervention would close it
Large unexplained share	Gap not captured by the three factors	May reflect omitted variables, not discrimination or any specific cause
Indirect effect interval excludes 0	Consistent with the literacy, self-efficacy, engagement pathway	Cross-sectional; order cannot be established
Alpha below 0.70 for a dimension	Scores for that dimension are unreliable	Revise items before using the dimension

Expected outcomes of the full study. Once the data are analysed the study is intended to yield the deliverables below. They are intended outputs, not findings, and each depends on the data supporting it.

TABLE XIV
EXPECTED OUTCOMES AND THE CONDITIONS THEY DEPEND ON

Outcome	What it would consist of	Condition
Empirical evidence	Estimated rural-urban differences with intervals for each dimension (RQ1, RQ2)	Adequate sample and response quality
Factor analysis	Explained and unexplained shares and the pathway estimate (RQ3, RQ4)	Reliable instruments; tenable model assumptions
Validated metrics	Reliability and validity evidence for the DLEQ, DLA and IAI, with revised items	Pilot meets thresholds; factor analysis added
Policy directions	Hypotheses for resource allocation, framed as suggestions to test	A gap with an identifiable associated factor
Intervention directions	Candidate actions for educators and institutions	Not supported by this design alone; needs an intervention study

IX. DISCUSSION, LIMITATIONS AND THREATS TO VALIDITY

The paper's contribution is a specified and tested analysis framework and a protocol. Whether the framework yields insight depends on data not yet analysed here, and nothing in this paper bears on whether rural students in fact fare worse. The limitations below are those we can already identify.

Practical implications, conditional on future results. The framework is designed so that either outcome is informative. If a gap in the DLE index is found and a large part is associated with infrastructure, resource allocation toward connectivity and devices would be the first candidate for follow-up, which is consistent with the emphasis in [11]. If the gap is mostly associated with literacy and self-efficacy, training-focused interventions would be indicated, in line with [12], [13]. If no gap is found, as in [8], [9], the measured dimensions may not separate the groups in this population, and that would still be useful to policy discussions that assume a large divide. In every case the design supports hypotheses for intervention studies, not intervention claims. It also bears on SDG 4: equitable quality in digital education can only be monitored if outcomes are measured, not just access.

Positioning against the reviewed work. Compared with single-dimension studies [3], [8], [9], DLEF measures several dimensions at once. Compared with outcome studies in other countries [5], [6], [7], it targets Indian higher education in one district. Compared with studies that explain gaps by regression or mediation [5], [11], [12], [13], it adds an explicit decomposition of the gap. These differences are of design and scope. None is shown here to give better estimates than existing approaches, and the novelty claim rests on the combination, which a systematic review should still check.

Evidence base. The literature review covers 15 studies, read at abstract level, retrieved through one search service. It is not systematic, and the statement in Section II-F that no reviewed study combines the compared elements cannot be read as proof of novelty. A systematic search of several databases, with full-text reading and a recorded search protocol, is needed. Several cited studies are from 2026 and have no citation history yet.

Construct validity. The three instruments are new and unvalidated. Alpha shows internal consistency, not dimensionality, and no confirmatory factor analysis is implemented. The DLE index is a standardised composite with equal weights, so it is sample-relative and its meaning depends on the weights. Self-reported performance impact may differ from actual performance, which is why academic records are collected as a check.

Omitted factors. Teachers' preparedness for digital instruction, which [7] and [11] identify as important, is not measured, nor is family involvement [7]. Their contribution would sit in the unexplained part of the decomposition, so a large unexplained share must not be read as evidence against the factors that were measured.

Internal validity. The design is cross-sectional, so the decomposition and mediation results are associational. The decomposition assumes linear relations and cannot separate omitted variables from true unexplained differences. Mediation cannot establish the temporal order of literacy, self-efficacy and engagement. Students are nested in institutes, and the current analysis treats them as independent. Institute-level clustering would make standard errors too small, and a multilevel or cluster-robust analysis is needed.

External validity. The population is one district of one state. If respondents come from the authors' networks, the sample is likely to over-represent students similar to the authors. A binary rural-urban split also hides peri-urban cases and students who live in one place and study in another.

Statistical conclusion validity. With 150-200 students per group only differences of about 0.28-0.32 standard deviations or larger are detectable at 80% power (Section IV-G). Subgroup analyses by level, gender or discipline will be underpowered. Likert scores are treated as interval, which is common practice but debatable, and the Mann-Whitney test is run alongside as a check. Missing data are handled by complete cases.

Implementation. The 15 tests use synthetic data with simple structure, so passing them does not rule out problems with real, messier data. The robustness and benchmark scripts have not been executed.

Summary of threats and planned mitigation.

TABLE XV
THREATS TO VALIDITY AND PLANNED MITIGATION

Threat	Type	Planned mitigation	Residual risk
Convenience or snowball sampling	External	Paper and online administration; report sample composition; describe results as sample-specific	Results cannot be generalised
Unvalidated instruments	Construct	Expert review, pilot, alpha, planned factor analysis	Dimensionality untested
Self-report bias	Measurement	Academic records as a check; anonymity	Shared-method variance remains
Clustering by institute	Statistical	Estimate intraclass correlation in the pilot; cluster-robust errors planned	Not yet implemented
Cross-sectional design	Internal	Associational wording; no causal claims	Order of variables unknown
Omitted factors	Internal	Cautious reading of the unexplained share	Teacher preparedness unmeasured
Low power for small effects	Statistical	Power table; report intervals	Subgroup analyses underpowered

X. FUTURE WORK AND CONCLUSION

Next steps, in order. (1) Confirm ethics clearance and the actual sampling method. (2) Run the pilot and revise the instruments. (3) Collect the main data and run the pipeline once, with the analysis plan fixed in advance. (4) Add the missing components: confirmatory factor analysis, cluster-robust or multilevel estimation, the Spearman validation, and the robustness and benchmark runs on real data. (5) Replace the abstract-level literature review with a systematic, full-text one. (6) Consider an intervention component, which would address gap G4, only after the descriptive picture is known.

Conclusion. We specified DLEF, a framework for comparing digital learning effectiveness between rural and urban higher-education students on six dimensions in one sample, and gave a protocol, a mathematical formulation and a tested reference implementation. The framework combines established statistical methods and proposes the composite DLE index and a specific application of the Blinder-Oaxaca and mediation analyses to this question. The paper reports no empirical findings. Whether rural students experience lower digital learning effectiveness than urban students, and how much of any difference is associated with infrastructure, literacy and self-efficacy, remains for the data to say.

DATA AND CODE AVAILABILITY

The reference implementation, with its 15-test suite, pinned dependencies, the data dictionary and the collection template, is available from the corresponding author on reasonable request, and the authors intend to deposit it in a public repository with a persistent identifier on acceptance. The survey responses contain information about identifiable individuals' education and location, so the respondent-level dataset will not be made public. Aggregate result tables and the run metadata will be shared, and respondent-level data may be shared in de-identified form only where ethics approval and participant consent allow. No dataset was analysed in this version of the manuscript, and the synthetic data used for testing is not study data.

APPENDIX A. RUNNING THE REFERENCE IMPLEMENTATION

The analysis on collected data is run from the command line on a CSV exported from the main-survey sheet of the collection template (same column names as the data dictionary):

```
python3 run_study.py responses.csv \
    --out results/ --seed 0 --n-boot 5000
```

The runner writes the reliability, group-comparison, decomposition and mediation tables and a metadata file, and prints no interpretation. The test suite is run with `python3 -m pytest tests`. The synthetic generator exists only to test the code and is labelled in every file it produces. Sensitivity and benchmark scripts (`robustness.py`, `benchmark.py`) are provided and, as stated in Section VI, have not been executed on study data.

REFERENCES

[1] T. Muthuprasad et al., "Students' perception and preference for online education in India during COVID-19 pandemic," *Social Sciences & Humanities Open*, vol. 3, no. 1, 2021.

[2] A. Lokhande et al., "Digital Education in India: Bridging the Divide or Widening the Gap?," *Int. J. Research in Applied Science and Engineering Technology*, 2026.

[3] B. T. Sampath Kumar et al., "The digital divide in India: use and non-use of ICT by rural and urban students," *World J. Science, Technology and Sustainable Development*, 2018.

[4] A. R. Sathyan et al., "Digital competence of higher education learners in the context of COVID-19 triggered online learning," *Social Sciences & Humanities Open*, 2022.

[5] L. Zhao et al., "Determinants of the digital outcome divide in e-learning between rural and urban students: Empirical evidence from the COVID-19 pandemic based on capital theory," *Computers in Human Behavior*, 2021.

[6] M. Raza et al., "A Comparative Study of Digital Divide and Learning Outcomes between Urban and Rural Undergraduates," *Social Prism*, 2026.

[7] H.-Y. Wang et al., "Revisiting the Digital Divide in Online Teaching: An Analysis of Urban-Rural Differences in Online Learning among Middle School Students," *Open Journal of Social Sciences*, 2022.

[8] T. Kavitha, "Digital Learning Preparedness and Technology Engagement among Higher Education Students: An Empirical Study," *Int. J. Informative and Futuristic Research*, 2026.

[9] K. Dhanasree et al., "Exploring Digital Competencies and Online Learning Readiness among Undergraduate Students," *Advances in Research*, 2025.

[10] F. Bast, "Perception of Online Learning Among Students From India Set Against the Pandemic," 2021.

[11] G. Tiwari et al., "Digital Divide and Its Impact on English Language Learning Among Rural vs. Urban ESL Learners in India," *Int. J. Humanities, Commerce and Education*, 2026.

[12] S. Getenet et al., "Students' digital technology attitude, literacy and self-efficacy and their effect on online learning engagement," *Int. J. Educational Technology in Higher Education*, 2024.

[13] Y. Feng et al., "Digital literacy is associated with college students' learning engagement via sequential mediation of self-efficacy and learning adaptability," *Scientific Reports*, 2026.

[14] S. Bandyopadhyay et al., "Exploring Rural-Urban Education Divide in India," book chapter, 2021.

[15] A. Mathrani et al., "Rural-Urban, Gender, and Digital Divides during the COVID-19 Lockdown: A Multi-Layered Study," *Societies*, vol. 13, no. 5, art. 122, 2023.

[16] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297-334, 1951.

[17] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Statist.*, vol. 18, no. 1, pp. 50-60, 1947.

[18] L. V. Hedges, "Distribution theory for Glass's estimator of effect size and related estimators," *J. Educ. Statist.*, vol. 6, no. 2, pp. 107-128, 1981.

[19] B. Efron, "Bootstrap methods: Another look at the jackknife," *Ann. Statist.*, vol. 7, no. 1, pp. 1-26, 1979.

[20] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Statist.*, vol. 6, no. 2, pp. 65-70, 1979.

[21] R. L. Oaxaca, "Male-female wage differentials in urban labor markets," *Int. Econ. Rev.*, vol. 14, no. 3, pp. 693-709, 1973.

[22] A. S. Blinder, "Wage discrimination: Reduced form and structural estimates," *J. Human Resources*, vol. 8, no. 4, pp. 436-455, 1973.

[23] D. Neumark, "Employers' discriminatory behavior and the estimation of wage discrimination," *J. Human Resources*, vol. 23, no. 3, pp. 279-295, 1988.

[24] K. J. Preacher and A. F. Hayes, "Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models," *Behav. Res. Methods*, vol. 40, no. 3, pp. 879-891, 2008.