Decomposing ethnic achievement gaps across multiple levels of analysis and multiple ethnic groups: A mediation approach

Beatriz Gallo Cordoba[1], George Leckie[2], William Browne[2]
[1]Faculty of Education, Monash University, Australia, [2]Centre for Multilevel Modelling, University of Bristol, United Kingdom

Abstract

A long-standing problem in the ethnic achievement gap literature is whether the gaps can be attributed to students or schools. In an attempt to solve this problem, three approaches have been proposed to decompose the Black-White achievement gap into components that can be attributed to students and schools. Such approaches have been limited because they do not consider additional levels of the education system (e.g. local authorities) or multiple ethnic groups (e.g. White, Black, Hispanic, Asian).

These three decomposition approaches fit into the mediation analysis framework. Therefore, we propose using this framework to extend the current decomposition approaches to consider additional levels of the education system and multiple ethnic groups.

Such extensions allow for more detailed insights about ethnic achievement gaps that can improve the pertinence of policy and future research recommendations. We illustrate the importance of such extensions using an application to data for Colombian students.

Keywords: Applied statistics; Mediation Analysis; Regression; Gap Decomposition.

E-mails: Beatriz.GalloCordoba@monash.edu; g.leckie@bristol.ac.uk; William.Browne@bristol.ac.uk

1. Introduction

Ethnic achievement gaps, usually defined as the difference between the mean test scores of two groups of students (e.g. Black and White) are usually attributed to differences in a wide range of variables defined at the student, school and higher levels of the education system (Bidwell & Kasarda, 1975; Rothstein, 2004). Within the context of studying the Black-White achievement gap in the US, three competing methodological approaches have been proposed to attribute this gap to differences between students or schools. We refer to these as Approach 1 (Cook & Evans, 2000; Fryer & Levitt, 2004, 2006), Approach 2 (Hanushek & Rivkin, 2006) and Approach 3 (Page et al., 2008; Reardon, 2008). Authors applying Approaches 1 and 3 argued that between-school differences explain at most 40% of the White-Black achievement gap in the US, while authors applying Approach 2 argued that these differences explain 70% of the achievement gap.

These current approaches are limited in at least two ways. First, they do not consider the role of other levels of the school system, such as school districts (districts). Second, they are restricted to the binary White-Black comparison, ignoring the role of other minority groups. These omissions to features of real data may hinder the pertinence and relevance for policy of such a gap decomposition analysis.

In this presentation we propose using mediation analysis as a framework to extend the three current decomposition approaches to consider multiple ethnic groups (e.g. White, Black, Asian, Hispanic) and additional levels of the school system (e.g. districts). The reader should note that we do not use mediation analysis in the traditional sense, but as a statistical tool for effect decomposition (Hou, 2014). The use of mediation analysis as a methodological tool to extend the ethnic achievement gap decomposition approaches to multiple levels and multiple groups has not been previously considered in the literature.

Section 2 presents the existing decomposition approaches under the mediation analysis framework, section 3 shows how this framework can be used to extend these approaches to multiple levels and multiple ethnic groups, section 4 presents the results of applying these three approaches to study ethnic achievement gaps in Colombia and section 5 concludes.

2. Current Approaches for the Within and Between-School Gap Decomposition

Reardon (2008) and Page et al. (2008) have previously reviewed the three existing methodological approaches for two-level two-group ethnic achievement gap decomposition and discussed their interpretation. This section summarizes these approaches from a mediation analysis perspective, in preparation for the multiple-level multigroup extension we present as our contribution in section 3.

Consider the single-level linear regression model

$$y_{ij} = \alpha + \beta M_{ij} + e_{y_{ij}}$$
$$e_{y_{ij}} \overset{\text{i.d.d}}{\sim} (0, \sigma^2) \tag{1}$$

where the dependent variable y_{ij} is the test score of student i attending school j and the independent variable M_{ij} is a dummy variable that equals 1 for ethnic minority students and 0 otherwise. In this case, β represents the average difference in test scores between minority and White students; the overall ethnic achievement gap.

Differences between students attending the same school and schools with different ethnic composition can be estimated using the contextual or hybrid-effect model (Mundlak, 1978)

$$
\begin{aligned}
y_{ij} &= \alpha + \beta^W(M_{ij} - \bar{M}_{.j}) + \beta^S \bar{M}_{.j} + e_{y_{ij}} \\
&= \alpha + \beta^W M_{ij} + (\beta^S - \beta^W)\bar{M}_{.j} + e_{y_{ij}} \\
&= \alpha + \beta^W M_{ij} + \beta^C \bar{M}_{.j} + e_{y_{ij}}
\end{aligned} \tag{2}
$$

where $\bar{M}_{.j}$ is the proportion of minority students in school j. Thus, β^W represents the within-school achievement gap; the average difference in test scores between white and minority students attending the same school. Additionally, β^S represents the between-school gap; the average difference in test scores between schools that only serve White students ($\bar{M}_{.j} = 0$) and schools that only serve minority students ($\bar{M}_{.j} = 1$). The last line in (2) illustrates the equivalence between the hybrid-effect model and the contextual effect model. In this last type of model $\beta^C = \beta^S - \beta^W$ represents the school contextual effect of ethnicity; the effect of attending a school with a larger proportion of minority students, over and above the effect of the students' own ethnicity.

To translate the three achievement gap decomposition approaches into the mediation analysis framework, it is enough to consider a model in which the ethnic composition of each school $\bar{M}_{.j}$ mediates the relationship between ethnicity M_{ij} and achievement y_{ij}, as we show in Figure 1.

The upper part of Figure 1 describes the total-effect model (1). The lower part of Figure 1 represents the outcome and mediation models. The outcome model is the contextual effect model (2). Besides, the mediation equation is

$$\bar{M}_{.j} = \gamma + \omega M_{ij} + e_{s_{.j}}, \forall i \in j \tag{3}$$

where the school proportion of minority students $\bar{M}_{.j}$ is repeated for all students i in school j. Here, γ represents the average school proportion of minority students experienced by White students, and ω represents the difference in the average school proportion of minority students experienced by White and minority students. If $\omega = 0$, White and minority students attend schools with the same proportion of minority students, while if $\omega = 1$, White students only attend schools with White students and minority students only attend schools with

minority students. Model (3) is a simple linear regression model that is used as a device to compute ω. Given its interpretation, ω is equivalent to the variance ratio index of segregation (Reardon, 2008).
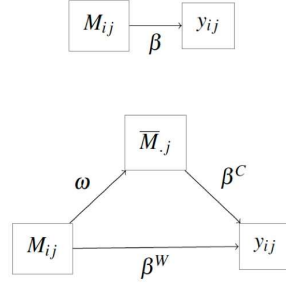

Figure 1: Model for the achievement gap decomposition into its within- and between-school components

In the mediation analysis context, the total effect β is decomposed into a direct effect β^W and an indirect effect $\omega\beta^C$, such that

$$\beta = \beta^W + \omega\beta^C \qquad (4)$$

Here, the direct effect is the within-school gap β^W and the indirect effect $\omega\beta^C$ represents how much more the test scores of minority students are affected by the school contextual effect of ethnicity β^C, in comparison to White students, given their average additional exposure to minority students ω. Reparametrising β^C in (4) leads to the different approaches that have been proposed for the decomposition of the ethnic achievement gap, as summarised in Table 1.

Table 1 Summary of current approaches for the ethnic achievement gap β decomposition into its student and school components.

	Approach 1	Approach 2	Approach 3
C1	β^W	$(1-\omega)\beta^W$	$(1-\omega)\beta^W$
C2	$\omega\beta^C$	$\omega\beta^S$	$\omega\beta^C$
C3			$\omega\beta^W$

The interpretation of the decomposition components in each of the approaches merits further discussion. However, our interpretation is that Approach 1 decomposes the overall gap β into the within-school achievement gap (the average differences between White and Minority students attending the same school) (C1) and the effect of segregation (C2); Approach 2 decomposes the gap into its within- (C1) and between-school (C2) components; and Approach 3 decomposes the gap into the within-school component of the gap (C1) and the effect of segregation through the school contextual effect of ethnicity (C2) and through differences in the schools' student intake (C3).

3. Considering Multiple Levels and Ethnic Groups

A generalised version of the mediation model that allows decomposing the overall gap of G groups (e.g. $G = 4$; White, Black, Hispanic, Asian) across L levels (e.g. $L = 3$; students, schools and districts) is shown in Figure 2 and Figure 3. For simplicity we omit the ij subscripts used previsously. This is a model in which the ethnic composition at each level l mediates the relationship between each ethnic group g and academic achievement y.

In this general case, the groups (categories of the independent variable) are indexed by $g = 1, \dots, G$ (where 1 is taken as the reference group), and the levels are indexed by $l = 1, \dots, L$. The dependent variable y and the dummy variables for each ethinc group D_g always vary by individual (level 1) and the mediators vary at level l. The coefficients' subscripts indicate effects that the group g 'receives' (e.g the overall gap β_g), while superscripts indicate effects that the group 'produces' (e.g. the contextual effect of group g at level l β^{lg}).
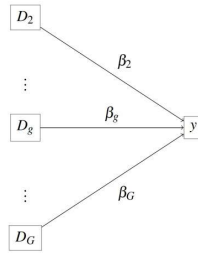

Figure 2: Total effect model for the achievement gap decomposition for G groups and L levels

The total effect model allows us to estimate the overall ethnic achievement gap β_g for each ethnic group, with respect to the reference category (e.g. the Black-White achievement gap)

$$y = \alpha + \sum_{g=2}^{G} \beta_g D_g + e_y \tag{5}$$
$$e_y \overset{\text{i.d.d}}{\sim} (0, \sigma^2)$$

The outcome model is equivalent to the contextual effect model

$$y = \alpha + \sum_{g=2}^{G} \beta_g^W D_g + \sum_{l=1}^{L} \sum_{g=2}^{G} \beta^{lg} E\left(D_g | l = l\right) + e_y \tag{6}$$

where β_g^W is the within-level-2 (and all upper levels) achievement gap between group g and group 1 and β^{lg} is the level l contextual effect of group g (e.g. the effect of attending a district with a higher proportion of Black students, over and above the school ethnic composition and the student's own ethnicity). The between-level-l achievement gap for group g is $\beta^{Blg} = \beta_g^W + \sum_l^L \beta^{lg}$.
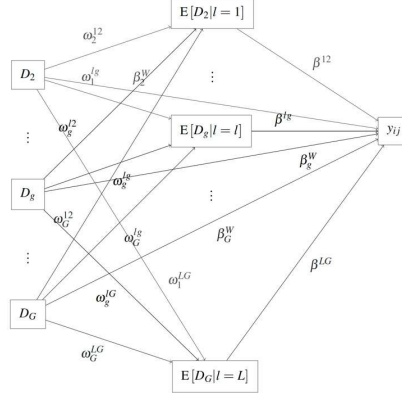
Figure 3: Outcome and mediation models for G groups and L levels

Since the model includes $(G-1) \times L$ mediators, there are $(G-1) \times L$ mediation equations given by

$$E(D_g | l = l) = \gamma_{lg} + \sum_{g'=2}^{G} \omega_{g'}^{lg} D_{g'} + e_{lg} \qquad (7)$$

Again, (6) and (7) are used as devices to decompose the ethnic achievement gap. In this case, (7) allows us to estimate the segregation indices ω_g^{lg}. In the mediation analysis framework (Hayes & Preacher, 2014), given these relationships, the overall gaps β_g can be decomposed as

$$\beta_g = \beta_g^W + \sum_{l=1}^{L} \sum_{g'} \omega_g^{lg'} \beta^{lg'}, g' = 2, \dots, G \qquad (8)$$

Again, reparametrizing β^{CS} and β^{CD} leads to the different decomposition approaches dicussed in section 2. These potential decompositions are shown in Table 2, where all the parameters are estimated using models (6) and (7). The expressions in this table highlight the fact that the gap for one of the groups (e.g. the Black-White achievement gap) cannot be decomposed without recognising the contextual effects of other ethnic groups (e.g. Hispanic and Asian). Additionally, C2 and C3 could be further decomposed to acknowledge the effect of each level l and group g when relevant.

Table 2 Achievement gap decomposition for group g when considering multiple levels and groups.

	Approach 1	Approach 2	Approach 3
C1	β_g^W	$(1 - \omega_g^{1g})\beta_g^W$	$(1 - \omega_g^{1g})\beta_g^W$

	Approach 1	Approach 2	Approach 3
C2	$\displaystyle\sum_{l=1}^{L}\sum_{g'} \omega_g^{lg'}\beta^{lg'}$	$\displaystyle\sum_{l=2}^{L}\left(\omega_g^{(l-1)g}-\omega_g^{lg}\right)\beta^{B(l-1)g}$ $+\omega_g^{Lg}\beta^{BLg}$	$\displaystyle\sum_{l=1}^{L}\sum_{g'} \omega_g^{lg'}\beta^{lg'}$
C3		$\displaystyle\sum_{l=1}^{L}\sum_{g'\neq g} \omega_g^{lg'}\beta^{lg'}$	$\displaystyle\sum_{l=2}^{L}\left(\omega_g^{(l-1)g}-\omega_g^{lg}\right)\beta_g^{W}$ $+\omega_g^{Lg}\beta_g^{W}$

3.1. Application

We apply the three existing decomposition approaches and our extended version to Colombian census SABER 11 data for the maths achievement of last year (11ᵗʰ grade, age 16/17) high-school students. The existing approaches decompose the White-Minority gap of 0.46 standard deviations (SD) into student and school components. In turn, our extended approach separately examines the gap between White and Afrocolombian (0.59 SD), Indigenous (0.52 SD) and other minority (0.14 SD) students, at the student, school and district levels. The result of these richer multiple-level multigroup decompositions are shown in Figure 4.
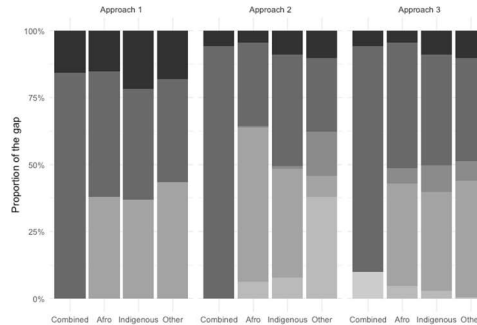

Figure 4: Application of the existing and extended decomposition approaches for the ethnic achievement gap in maths scores for 11th Grade Colombian Students. 'Combined' represents the existing approaches
Source: SABER 11, 2011. Icfes.

4. Conclusions

Each of the three traditiaonl decomposition approaches and the extentsions we have proposed for multilevel multigroup settings is useful to inform different research questions and policy decisions. For example, if the debate is about school segregation, Approach 1 provides a more direct way to analyse its potential effects on the achievement gap. If the focus is on the within-school and between school components of the gap, Approach 2

decomposition is appropriate. If, in turn, the interest is in the mechanisms behind the between-school component of the gap, Approach 3 decomposition can be used.

The application to Colombian data shows the potential of our extensions to transform the discussion around ethnic achievement gaps. Considering districts in the decomposition showed they play a role at least as important as schools and thus the need of policies to consider them. Similarly, the heterogeneity among ethnic minority groups that is ignored when treating them as a single group shows the importance of extending the approaches to consider multiple ethnic groups.

References

Bidwell, C. E., & Kasarda, J. D. (1975). School District Organization and Student Achievement. *American Sociological Review*, *40*(1), 55. https://doi.org/10.2307/2094447

Cook, M. D., & Evans, W. N. (2000). Families or Schools? Explaining the Convergence in White and Black Academic Performance. *Journal of Labor Economics*, *18*(4), 729–754. https://doi.org/10.1086/209975

Fryer, R. G., & Levitt, S. D. (2004). Understanding the Black-White Test Score Gap in the First Two Years of School. *The Review of Economics and Statistics*, *86*(May), 447–464. https://doi.org/10.1162/003465304323031049

Fryer, R. G., & Levitt, S. D. (2006). The Black-White test score gap through third grade. *American Law and Economics Review*, *8*(2), 249–281. https://doi.org/10.1093/aler/ahl003

Hanushek, E. A., & Rivkin, S. G. (2006). School quality and the black-white achievement gap. In *NBER Working Paper Series* (No. 12651; Issue October). http://www.nber.org/papers/w12651

Hayes, A. F., & Preacher, K. J. (2014). Statistical mediation analysis with a multicategorical independent variable. *British Journal of Mathematical and Statistical Psychology*, *67*(3), 451–470. https://doi.org/10.1111/bmsp.12028

Hou, F. (2014). A General Approach to Effect Decomposition. *Social Science Quarterly*, *95*(3), 894–904. https://doi.org/10.1111/ssqu.12074

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica*, *46*(1), 69–85. http://www.jstor.org/stable/1913646

Page, L., Murnane, R., & Willett, J. (2008). *Trends in the Black-White Achievement Gap: Clarifying the Meaning of Within- and Between-School Achievement Gaps* (No. 14213). https://doi.org/10.3386/w14213

Reardon, S. F. (2008). *Thirteen Ways of Looking at the Black-White Test Score Gap* (No. 2008–08; Significant Education Policy and Practice Issues). http://web.stanford.edu/group/cepa/workingpapers/WORKING_PAPER_2008_08.pdf

Rothstein, R. (2004). A Wider Lens on the Black-White Achievement Gap. *Phi Delta Kappan*, *86*(2), 104–110. https://doi.org/10.1177/003172170408600204