AN EMPIRICAL EVALUATION OF THE DISAGGREGATED EFFECTS OF
EDUCATIONAL DIVERSITY IN A NATIONAL SAMPLE OF LAW SCHOOLS

Nisha C. Gottfredson

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Psychology.

Chapel Hill
2008

Approved by:
A. T. Panter, Ph.D.
Daniel J. Bauer, Ph.D.
Patrick J. Curran, Ph.D.

ABSTRACT

Nisha Gottfredson: An empirical evaluation of the disaggregated effects of educational
diversity in a national sample of law schools
(Under the direction of A T. Panter, Ph.D)

The use of race-conscious admissions practices to achieve student diversity in

academic institutions has recently been challenged. An understanding of how racial diversity

in law school affects students is useful to develop administrative policies that support social

and intellectual growth of students after they are admitted. A nationally-representative

sample of 2,180 students from 64 accredited U.S. law schools was used to model the

mechanism through which institutional diversity may influence student outcomes in a

multigroup, multilevel SEM framework. Results suggest that racial heterogeneity directly

and indirectly increases exchange of ideas and decreases racist/classist attitudes. The effects

of racial diversity were mediated by increased contact with racially diverse peers. Results

were similar for White and non-White students. This study confirms the usefulness of

admissions policies that permit racial diversity in academic institutions, and imply that

educators should focus on increasing intergroup contact between students.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

An Empirical Evaluation of the Disaggregated Effects of Educational Diversity in a National

Sample of Law Schools

Opinion and practices relating to racial equality in education have changed

dramatically since the *Brown v. Board of Education* ruling in 1954. Between the 1960s and

the 1980s, the many institutions focused on affirmative action programs that would promote

racial equality by fostering assimilation and integration of racial and ethnic minorities into

mainstream White culture (Edwards, 2004). At the same time, discomfort with the idea of

equalizing access to higher education through affirmative action is evident in a number of

Supreme Court cases that involved White plaintiffs arguing against racial quotas or the

consideration of race in admissions into institutions of higher education (e.g., *Regents of the*

University of California v. Bakke, 1978; *Gratz v. Bollinge*r, 2003; *Grutter v. Bollinger*,

2003). Meanwhile, resistance to the idea of assimilation into White culture at the expense of

heritage and culture has also become a point of contention for many minority groups

(Edwards, 2004). Both of these factors, along with research findings that suggested a variety

of benefits that arise from diverse learning environments (e.g., Antonio, 2001; Antonio,

Chang, Hakuta, Kenny, Levin & Milem, 2004; Bowen & Bok, 1998; Chang, Denson, Sáenz

& Misa, 2006; Gurin, Dey, Hurtado & Gurin, 2002; Holzer & Neumark, 2006; Niemann &

Maruyama, 2005) have redirected the emphasis of race-conscious admissions away from

equalizing access *per se* and towards the potentially instrumental value of educational

diversity and multiculturalism for producing other desirable student outcomes.

A notable example of this trend occurred in 2003 when the Supreme Court ruled that

institutions of higher education have a "compelling interest in attaining a diverse student

body" (*Grutter, p.* 16). The case dealt with a lawsuit against the University of Michigan Law

School and was brought by Grutter, a White student who claimed that she was being

discriminated against based on her race when she was not admitted to the law school. The

Court, in a 5-4 majority, favored the Law School. It reasoned that a diverse student body

could promote the mission of the Law School to train future citizens and leaders to become

successful in an increasingly diverse society. The Court ruled that it is within the

constitutional rights of academic institutions to diversify their student bodies, as long as race

is considered a subjective "plus factor" in the decision-making process. This approach is in

contrast to the quantifiable approach used by undergraduate admissions at the University of

Michigan (*Gratz v. Bollinger,* 2003) that is now illegal because of its explicit consideration

of race in admissions.

Despite ruling in favor of the Law School and its race-conscious admissions policies,

the Court majority implied that at some point (perhaps 25 years in the future) race-conscious

admissions practices may no longer be appropriate. Recent Supreme Court rulings under the

recently appointed Chief Justice Roberts have suggested that an even earlier prohibition of

race-conscious admissions policies may be inevitable (e.g., *Parents v. Seattle*, 2007;

Meredith v. Jefferson Co. Board of Education, 2007).

In a number of states (including California, Michigan, Texas, and Washington), race-

neutral admissions policies have been required in response to state-level legislation banning

the use of race in admissions decisions. One example of state legislation was California's Proposition 209, which banned the consideration of demographic factors in admissions and severely reduced minority representation at Berkeley and UCLA (Allen, Jayakumar, Griffin, Korn, & Hurtado, 2006). Some proposed race-neutral admissions plans include automatic admission for the top 10% of high school (or college) students from every graduating class or random admissions lotteries among applicants. These suggested approaches raise a number of problems, as illustrated by the Texas "Ten Percent Plan" (e.g., reducing minority representation at flagship schools but increasing enrollment of less qualified Whites at these schools; Bucks, 2003). In addition, these approaches are inferior to the consideration of race in admissions if the aim of an educational institution is to construct an ethnically/racially diverse *and* academically talented class of students. Holzer and Neumark (2006) have recommended following the lead of businesses that aim to employ diverse workforces. These businesses recruit and screen candidates more extensively, provide more training, and carefully evaluate worker performance. Such policies prevent businesses from lowering employee standards while increasing workplace diversity.

A clever suggestion for crafting a diversity-conscious, race-neutral admissions formula for law schools has recently been proposed by Shultz (2007). Shultz conducted focus groups with lawyers, judges, professors and students to identify the skills necessary for success in the legal field. She identified 26 "effectiveness factors" that can be assessed in prospective law students to predict whether they will become effective lawyers. The consideration of factors such as personality to predict future performance has been widely used in the domain of business, but remains untested in educational admissions. While Shultz's (2007) proposed plan is more time and energy intensive for admissions committees

3

than the reliance on a smaller number of admissions criteria, this suggestion may be the most promising to date.

Prior to the *Grutter* decision in 2003, LSAC funded researchers from the University of North Carolina at Chapel Hill and Greensboro and from the University of California at Los Angeles to form the Educational Diversity Project (EDP). One of the main goals of this project is to determine how much, if at all, diversity, including racial diversity, affects student outcomes, and if so, to understand the mechanisms through which this occurs. Such understanding will assist administrators in maximizing educational outcomes for their students in the face of reduced flexibility in admissions criteria.

In this investigation, I will examine whether there are measurable educational effects of diversity, with a focus on understanding ethnic/racial diversity. I adopt this focus because it is the only type of educational diversity that is currently being challenged in courts. The relationships to be tested involve aspects of student diversity that are presumed to be directly linked to a racially-diverse class of students, yet can be directly manipulated by school administrators should race-conscious admissions become more difficult to employ due to legal obstacles. The following paragraphs elaborate on social and psychological research and legal precedents that bear on this research objective.

Prejudice reduction through intergroup contact. The University of Michigan Law School strives to accept "a mix of students with varying backgrounds and experiences who will respect and learn from each other" (*Grutter v. Bollinger*, 2003, p.1). The Law School argued, and the Court concurred, that a critical mass of historically- underrepresented and discriminated against ethnic/racial groups must be enrolled for minority students to participate in class without feeling isolated or like "spokespersons" for their race or ethnic

4

group. Kent Syverud, a former University of Michigan Law School professor, testified during *Grutter* that the presence of a critical mass of underrepresented students causes a reduction in racial stereotypes. He opined that the critical mass helps students to understand that there is "no minority viewpoint, but rather a variety of viewpoints among minority students" (*p.* 7). In essence, he testified that student prejudices about racial minorities at law school are reduced in the presence of a diverse student body.

In 1954, Allport laid out optimal conditions for reducing prejudice in environments containing diverse groups of people, a perspective that has come to be known as contact theory. According to Allport, mere contact between groups is not sufficient to reduce prejudice because "prejudice screens and interprets our perceptions" (*p.* 252). True knowledge and understanding of so-called "outgroups" are required for prejudice reduction to occur. Allport theorized that not only must people be frequently exposed to members of an outgroup in casual situations, but also that groups should be of equivalent social standing in a cooperative environment, and the interactions should be supported and encouraged by an authority figure or an institution.

In a recent meta-analysis of 515 studies of contact theory, Pettigrew and Tropp (2006) confirmed that increased contact with outgroups, even when not under the optimal conditions specified by Allport, showed a consistent, small-to-medium effect in reducing prejudice and increasing positive attitudes about an outgroup. Further, Pettigrew and Tropp (2006) found that the benefits of frequent contact generalize beyond the specific outgroup of exposure (e.g., to other ethnic/racial groups).

Confronting criticism that research findings supporting contact theory might be caused by selection (e.g., people who are prejudiced avoid outgroups while non-prejudiced

people seek contact with outgroups), Powers and Ellison (1995) employed endogenous switching regression models, an econometric technique that compares causal models. They found no evidence of selection bias, providing support for the hypothesized directionality of outgroup contact on prejudice.

These findings form the basis for one of the diversity constructs developed in this study. *Intergroup contact* measures the frequency with which an individual interacts with peers of different ethnic/racial backgrounds. I hypothesize that frequent interactions with diverse others of equal social standing in academic environments results in general prejudice reduction as suggested by Allport (1954). Students attending the same law school meet Allport's condition of having equal social standing because they were both selected for admission and chose to attend the same law school. A study examining EDP's baseline sample, which modeled the impact of cooperative intergroup contact during undergraduate years on attitudes favoring ethnic/racial and socio-economic minorities, suggested that this type of frequent, casual, cooperative contact with ethnically/racially diverse peers diminishes prejudice (Gottfredson, Panter, Daye, Allen, Wightman & Deo, in press).

Outgroup prejudice is not specific to a single group of individuals. Contact theory suggests that integroup contact should reduce general prejudice against outgroups (Allport, 1954). It is for this reason that I am using the construct titled *perceptions of a meritocracy* to measure general outgroup prejudices. Researchers Federico and Sidanios (2002) have identified two main components that account for dominant group members' objections to policies such as affirmative action. The first, "principled conservatism," is linked to individualistic work ethic and politically conservative values. The second is "general group-dominance," a desire to maintain a privileged position at the expense of other groups. Reyna,

Korfmacher, and Tucker (2005) report that these two attitudes tend to co-occur; outgroup-based stereotypes may mediate the effect of conservatism on anti affirmative-action attitudes. After controlling for an individual's political orientation, measures of perceptions of a meritocracy remain predictive prejudice against economically and socially disadvantaged groups.

   Racial diversity and the exchange of ideas. In *Grutter*, the University of Michigan Law School argued that student learning is enhanced when students are surrounded by other students who do not share the same beliefs and values and who have not had the same life experiences. In a diverse student body each student contributes to the learning of other students by providing a different perspective on issues. In law school, this diverse exchange of ideas would mainly take place during classroom discussions and in other extracurricular academic settings.

   An exchange of diverse viewpoints among students of the same racial/ethnic background is possible to some degree because every student, regardless of their race/ethnicity, has had a different set of life experiences can provide a new perspective. As shown by Panter, Daye, Allen, and Wightman (2005), however, race/ethnicity accounts for a significant proportion of variation in a variety of life experiences, family and neighborhood histories, attitudes, beliefs, and values. Furthermore, as Lempert testified in the *Grutter* (2003) case, students who have experienced a lifetime of racial discrimination may be able to offer perspectives not available to most White students. Such perspectives are particularly important for aspiring lawyers who will encounter situations involving complicated racial/ethnic dynamics and who may enter a field of law that focuses on issues involving

underrepresented populations. This research forms the basis for another measure used in this study, *diversity of ideas*.

Objectives of the Research

The importance of student racial/ethnic diversity has been demonstrated by a number of scientists using different methodologies. Antonio et al. (2004), using racially diverse and homogenous focus groups as experimental conditions, demonstrated that complex thinking significantly increases when college students are surrounded by racially diverse peers. In addition to this experimental evidence, a number of correlational studies imply that racial diversity at educational institutions improves a number of student outcomes (Shaw, 2005) including increased citizenship, academic skills, and intellectual engagement (Gurin et al., 2002), increased cognitive openness (Gottfredson et al., in press), and increased cultural knowledge (Antonio, 2001). These findings suggest that there would be value in continuing race-conscious admissions practices in higher education, but may not be sufficiently persuasive to prevail in the contemporary legal climate.

Ethnic/racial diversity alone is not always sufficient to benefit student learning (e.g., Allport, 1954; Gurin et al., 2002; Holzer & Neumark, 2006; Niemann & Maruyama, 2005). Supportive environments also appear to be essential for harvesting the benefits of diversity. The results of the present study should provide those who make policy with knowledge of which types of diversity influence various student outcomes so that they can allocate limited educational resources to pursuing the most efficacious practices.

A number of race-neutral alternatives to admissions formulae have been proposed, some of which could potentially result in an academically talented and racially diverse student bodies that are on par with, or even superior to, those that could be obtained through

8

race-conscious admissions (e.g., Shultz, 2007 for law school admissions). Yet it is important also to model the precise mechanisms through which ethnic/racial diversity influences student outcomes. For example, what are the effects of diversity of ideas and intergroup contact on perceptions of a meritocracy? Does intergroup contact mediate the beneficial effects of racial heterogeneity? What, if any, benefits do the racial heterogeneity of a school provide above and beyond intergroup contact?

The present study utilizes a large nationally-representative sample of a law students and a large volunteer sample of law students. Hypothesized directional relationships are tested with data using a series of models that capture several features of the data. Baseline measures of attitudes and obtained as the law students entered law school in Fall 2004 and were reassessed when the students exited law school in Spring 2007, along with several additional assessments of experiences during law school. School-level measures of racial heterogeneity, sector, selectivity, and enrollment were obtained from publicly available data provided by the American Bar Association (ABA). Constructs that cannot be assumed to be measured without error were modeled as latent variables, and multiple-group models were tested.

This research was motivated by a practical question of contemporary importance. I attempt to model parsimoniously the structural mechanism by which educational diversity may influence educational outcomes and which potentially differs across ethnic/racial groups. The nesting of students within schools necessitates a multilevel modeling framework (e.g., Raudenbush & Bryk, 2002) while my inability to directly measure the constructs involved without measurement error dictates a latent variable modeling framework (e.g.,

Thurstone, 1947). In doing so, I demonstrate how to conduct a multigroup multilevel structural equation model (MSEM) with categorical indicators given the available resources.

Using appropriate statistical models for appropriate substantive questions has numerous benefits. In the present case, missing data, ordinal manifest variables, unobserved latent variables, and nesting of students within schools can all be modeled in the same state-of-the-art framework of MSEM. There are, however, inherent limitations to using newly developed or developing models. Accordingly, in addition to testing substantive theory, a goal of this manuscript is to contribute to continued development of the quantitative methods used by identifying some of the practical weaknesses of using MSEM with a real-world data situation.

Summary of Hypotheses

1. Contact diversity reduces racist/classist prejudice. There is strong support for the claim that intergroup contact, based on Allport's (1954) contract theory, reduces prejudice among members of different racial/ethnic groups. Here I will examine whether frequent, informal interactions with diverse peers reduces ethnic/racial and social class prejudice among law students.

2. Racial diversity promotes a diversity of ideas. Diversity of ideas in law school is measured in the present research by student reports of high quality class discussions, an atmosphere that allows a respectful exchange of diverse beliefs and political views, and a sense of openness to new ideas. I hypothesize that, after taking into account the attributes of the law schools that students choose to attend (enrollment, sector, and selectivity) and the student attributes (age, gender, LSAT score, SES, political orientation), more racially diverse schools will be characterized by greater diversity of ideas.

3. Intergroup contact mediates the effects of racial diversity. As suggested by a number of researchers (Allport, 1954; Gurin et al., 2002; Holzer & Neumark, 2006; Niemann & Maruyama, 2005), racial diversity has the potential for positive or negative effects on student outcomes. Benefits must be encouraged through cooperative contact between groups. Thus, the model specifies that intergroup contact mediates the effects of racial diversity on diversity of ideas within a school and on perceptions of a meritocracy.

4. Exchange of diverse ideas reduces prejudice. The final hypothesis is that a students' exposure to new and diverse ideas will reduce prejudice towards people of lower SES and minorities in that they will become disillusioned with previously-held beliefs that these groups have *earned* disadvantage through a meritorious system (e.g., because of their innate intellectual ability, work ethic, and so on). Theoretically, the effect of racial diversity on perceptions of a meritocracy might be partially mediated by increased exposure to alternative viewpoints. The sparse extant empirical literature on this topic now runs counter to the proposed hypothesis. Some studies have shown that students who enter school with unfavorable attitudes regarding affirmative action that coincide with high levels of racism tend to develop more socially acceptable "principled objections" to affirmative action as their education increases (Federico & Sidanious, 2002; Reyna, et al., 2005). The present research will add to the limited evidence bearing on the hypothesis.

Methodological Background

Muthén and Satorra (1995) discussed two approaches to modeling complex data in an SEM framework: aggregated disaggregated analyses. Aggregated analysis methods are useful for developing an overall population average model that does not generalize to any particular sampling unit. Disaggregated analysis estimates variability across independent

sampling units so that estimates generalize to individuals in a sample. Disaggregated analyses are more informative but they are more difficult to estimate and sensitive to misspecification. Both model types will be estimated in the course of the analysis with the final goal of estimating and interpreting parameters from disaggregated models.

Disaggregated multilevel structural equation models (MSEMs) were first developed in the late 1980s by Goldstein and McDonald (1988). In 1989, McDonald and Goldstein developed an analytical solution to obtain maximum-likelihood estimates for balanced and unbalanced MSEMs with continuous data; however, these solutions were impractical to implement in practice and were not widely available. In the same year, Muthén presented a maximum-likelihood based estimator (MUML) that could handle unbalanced data and that was computationally more feasible to implement than McDonald and Goldstein's (1989) true maximum-likelihood estimator. Unfortunately, MUML could not handle missing data or categorical outcomes. A number of researchers proposed estimation techniques for specific MSEM formulations throughout the 1990s and early 2000s (e.g., du Toit & du Toit, 2002; Lee & Poon, 1998; Raudenbush, 1995). In 2004, Liang and Bentler expanded Lee and Poon's (1998) technique by deriving an expectation-maximization (EM) algorithm for obtaining full information maximum-likelihood solutions with a generalized MSEM formulation. This relatively recent development has increased the feasibility of implementing MSEM analyses with missing data and has become incorporated into a number of software programs.

The presence of categorical data provides an added complexity to the estimation of MSEMs. The issue of non-continuous data in MSEMs was only recently confronted in the literature by Rabe-Hesketh, Skrondal, and Pickles (2004; also Skrondal & Rabe-Hesketh, 2004). When latent variable indicators are categorical, there is no closed-form solution to the

marginal likelihood (Rabe-Hesketh, Skronda, & Pickles, 2005). In this case, the likelihood

must be approximated by the computationally-intensive numerical integration procedure. The

most commonly used procedure for evaluating likelihoods of multilevel data with categorical

outcomes is Gauss-Hermite quadrature (Bock & Lieberman, 1970). Other techniques,

including adaptive quadrature, are also available. However, these techniques are even more

computationally intensive and are only necessary in the case of a large number of

independent sampling units (e.g., N=200) or in the case of large intra-class correlations (e.g.,

ICC=.35; Rabe-Hesketh, et al., 2005). MSEM for categorical data is now widely available in

software programs including M*plus* (Muthén & Muthén, 2007) and GLLAMM (Rabe-

Hesketh, Skrondal, & Pickles, 2006).

   As an alternative to full information maximum likelihood (FIML) estimation, a

diagonally weighted least squares statistic (WLSMV) that provides a standard error and

mean-adjusted chi-square test statistic is available in M*plus* (Muthén, du Toit, & Spisic,

1997; Muthén & Muthén, 2007; Wirth & Edwards, 2007). WLSMV depends on the analysis

of polychoric or tetrachoric correlation matrices rather than raw data. This limited

information estimator is less efficient than FIML, and requires often untenable assumptions

regarding missingness mechanisms (i.e., missing completely at random; Rubin, 1976).

However, there are two benefits to using the WLSMV estimator. First, computational time is

not dependent on the number of random effects in the model so it is more feasible for

estimating very complex models. Second, the two-stage estimation procedure enables the

estimation of a saturated model for the multilevel data, and is thus able to provide global fit

statistics that are not available with maximum likelihood estimation methods (Asparouhov &

Muthén, 2007).

MSEM has been frequently touted as an important procedure for modeling real-life psychological, sociological, and economic data given its ability to simultaneously handle nested data and data that has been measured with error simultaneously. A number of quantitative researchers have published expository papers demonstrating empirical examples of the use of the procedure (e.g., Kaplan & Elliott, 1997, Liang & Bentler, 2004; Muthén, 1994; Muthén, Khoo, & Gufstafsson, 1997; Skrondal &Rabe-Hesketh, 2004). However, these examples have involved idyllic data scenarios. For instance, most involved very simple CFA or MIMIC structures and all except the examples from Skrondal and Rabe-Hesketh (2004) predated the incorporation of software programming for declaring data as categorical (e.g., MSEM for categorical data was not implemented in M*plus* until 2004 in version 3).

Chapter 2

Method

Sample and Participants

National data were collected from law students at two time points. There were two
types of participants: those attending law schools selected in our core (nationally-
representative) sample and those attending law schools with administrators who volunteered
to participate in our study. Data were collected by a multisite, multidisciplinary team of
Educational Diversity Project (EDP) researchers during law school orientation in Fall 2004.
Student participants were again contacted in Spring 2007 as they were expected to be
completing law school. Web-based surveys were conducted through Qualtrics®, an online
survey software program. Institutional characteristics of the schools were provided by the
American Bar Association (ABA) on law schools. Table 1 presents descriptive statistics of
the volunteer and random law schools, and Table 2 provides information about the
demographic characteristics of students who participated in the study.

Participants in the volunteer sample include 1,963 incoming law school students from
16 ABA-approved, accredited law schools. Administrators of the 16 law schools in this
sample volunteered to be a part of EDP after hearing about the study through a presentation
at an annual Law School Admissions Council (LSAC) meeting, or after reading about the
study from a newsletter widely distributed to admissions counselors. Schools in the volunteer
sample are not a probability sample of law schools in the U.S., but they represent 12 states

spanning the continental United States and characteristics of students attending these schools do not differ significantly from those of student attending schools selected in the nationally representative sample. On average, 58.2% of law students per volunteer school completed baseline surveys.

Baseline participants in the core sample were 6,100 incoming law school students from 50 nationally representative, ABA-approved, accredited U. S. law schools. Schools in this sample were drawn using two methods. EDP investigators oversampled schools that were identified as having very high minority populations ($N = 7$), and randomly drew 46 schools from the remaining 177 ABA-approved law schools. Of these schools, one was ineligible to participate and two were non-responsive, resulting in a sample of 50 law schools. In the schools with high minority representations, average student response rates were 75.5%, and student response rates at the remaining schools were 51.8% on average. Higher response rates in the former are partially attributable to the administration method; all students in the high minority representation sample completed surveys during law school orientation, while the other sample included some students who completed the surveys during orientation and others who took surveys home with them.

In total, 8,063 students provided baseline data. Sixty-five percent ($N=5,258$) of the baseline participants consented to be re-contacted by EDP. Of these, 393 students could not be re-contacted due to invalid e-mail addresses. Out of the remaining participants, 2,695 either completed the follow up survey ($N = 2,180$) or were confirmed to have left law school prior to Spring 2007 ($N = 515$). Data from the volunteer and core samples were combined to obtain the maximum possible independent sampling units (ISUs) as necessary for obtaining stable estimates. This decision to combine the volunteer and core samples for the sake of

estimating a disaggregated model precluded the use of sampling weights that are available

for the core sample. The elimination of sampling weights in conjunction with the non-

random loss of individuals from baseline to follow-up (see below) limits my ability to

generalize study findings to law school students in the Nation.

Missing Data Considerations.

Taking into account the known reasons for attrition (permission to re-contact not

granted, invalid e-mail addresses, or law school dropout), 54.0% of nonresponders did not

complete the follow-up survey for reasons that are unknown to the EDP. A listwise or

pairwise deletion of cases for which follow-up data are missing might provide severely

biased results if the missingness mechanism is not completely at random (e.g., Little &

Rubin, 2002).

Table 3 compares baseline item responses of people who completed the follow-up

survey and those who failed to complete the follow-up survey. Attriters and non-attriters do

not appear to differ with respect to the amount of intergroup contact that they had during

college, or with respect to socioeconomic background during childhood. A higher proportion

of females and White students responded to the follow-up survey. Also, people who

responded to the follow-up survey were slightly less likely to be politically moderate or

slightly conservative, and were less likely to be neutral or to slightly agree with items

comprising the perceptions of a meritocracy construct at baseline (to be described later).

Results from inferential tests of group differences for these items are statistically significant,

but are generally uninformative about the practical differences given the large sample sizes in

both groups.

Fortunately, information about the most sensitive outcome variables was collected at baseline, along with a number of demographic measures. Controlling for these baseline variables, nonresponse to the follow-up items is not likely to be a function of unobserved data on the outcome variables of interest. The mechanisms of non-response (probably related to Bar Examination and job search-related stressors) should be uncorrelated with diversity experiences in law school. Thus, it is reasonable to assume that the follow-up data are missing at random (MAR; Little & Rubin, 2002), conditional on baseline covariates. If MAR assumptions are met, full information maximum-likelihood estimation techniques should provide unbiased parameter estimates by using all available data, including cases for which only partial information is available, rather than the sample covariance matrix in the likelihood function that only uses complete cases (e.g., Arbuckle, 1996; Enders & Bandalos, 2001; Wothke, 2000).

Full information maximum likelihood is not a magic bullet for handling missing data, and particularly not in this study. Parameters are estimated using a conditional likelihood of the endogenous variables on the exogenous variables. This means that cases for which follow-up data are not present are deleted from the analyses, greatly reducing statistical power to detect effects.

Multiple imputation (MI; Rubin, 1987) is also not an option for recovering missing data in this study. MI is a technique for obtaining unbiased parameter estimates and standard errors by creating a missing data model and predicting missing data values. Unlike other imputation techniques, MI incorporates random error that is consistent with the sampling error that would have existed had the values been observed. It is useful when data are MAR, but only when endogenous data are normal or dichotomous. Imputing categorical data results

18

in non-discrete values that must be either rounded, creating substantial bias (Horton, Lipsitz, & Parzen, 2003), or it must be treated as continuous, also introducing bias.

Measures

Institutional characteristics. Background attributes of the law school included: school enrollment, percent of applications accepted (the reverse of selectivity), and sector (*public* (0) or *private* (1)).

Background characteristics. Student self-reported demographics from the EDP baseline survey included: age, race/ethnicity (coded as White or non-White), gender, LSAT scores, political orientation (a five-point scale ranging from 1 (*Extremely Liberal*) to 5 (*Extremely Conservative*), and relative family household income during childhood (a five-point scale ranging from 1 (*Far Below Average*) to 5 (*Far Above Average*). Self-reported age, race/ethnicity, gender, and LSAT scores were verified with the LSAC databases.

Baseline and follow-up attitudes. On both the baseline and follow-up EDP survey, students answered questions pertaining to their socio-political beliefs on a Likert-type scale ranging from 1 (*Strongly Disagree*) to 5 (*Strongly Agree*). At both time points, students rated how much they agreed with the following statements: "In America today, every person has an equal opportunity to achieve success;" "Because Irish, Italians, Jews, and many other minorities overcame prejudice and worked their way up, Blacks should do the same without any special favors;" and "People at the bottom of the economic scale are probably lazier than those at the top." These variables comprise the perceptions of a meritocracy factor. Students who score highly on the perceptions of a meritocracy factor are inclined to attribute economic success and failure to individual characteristics, and to ignore social and historical barriers, or facilitators, to success. The follow-up measures are estimated in a measurement

19

model. However, due to the excessive computing time necessary for each additional random component estimated in a multilevel structural equation model with ordinal measures, the pre-test measures were summed to create a composite estimate of baseline attitudes.

Three diversities. Three types of diversity were measured in this study: (1) racial heterogeneity (Racial Diversity Index; *RDI*), (2) frequency of contact with diverse peers (intergroup contact), and (3) diversity of ideas. RDI is measured at the school level, and the other two diversities contain variance at both the individual and school level.

RDI, a measured variable, is calculated by summing the squares of the proportions of each ethnic/racial group in a school and subtracting that number from one (Formula 1; Lieberson, 1969). This creates a variable that is large (closer to one) when a school has many medium-sized ethnic/racial groups and is small (closer to zero) if there is a dominant ethnic/racial group:

$$RDI = 1 - \sum_{\substack{all\ ethnic \\ groups}} p^2 \qquad (1)$$

Intergroup contact is based on student reports of frequency of interactions with peers of various ethnic/racial backgrounds during college (measured at baseline) and law school (measured at follow-up). On a three-point scale from 1 (*never*) to 3 (*frequently*), students rated how often they interacted with African American, Asian American, Hispanic/Latino, Native American, and White peers. Frequencies were summed across the five groups to create an approximately normally distributed variable ranging from five to 15.

Lastly, diversity of ideas is a latent construct composed of four indicators. On a five-point Likert-type scale, students rated the quality of class discussions at their school, how much their school is characterized by respectful exchange of political views, how much their

school is characterized by respect for expression of diverse belief, and how open their school is to new ideas.

Data Analyses

Modeling ordinal latent variable indicators: Latent response formulation. As is typically the case in psychological measurement, latent constructs (perceptions of a meritocracy and diversity of ideas) were measured by asking participants to rate their beliefs and attitudes using a Likert-type ordinal response scale. In nonlinear factor analysis with ordered categories (i.e., a graded item response model; Samejima, 1969; Skrondal & Rabe-Hesketh, 2004; Wirth & Edwards, 2007), it is assumed that the item response choice is governed by how much an individual is characterized by the underlying, normally distributed common factor, *theta*. If an individual's level of theta falls below the lowest threshold, the participant will respond in the first category. If theta falls between the first and second thresholds, the participant will respond in the second category, and so on. The conditional probability that an individual will choose a particular response level can be modeled with a normal ogive, or, for computational ease, with a logit model. If a logistic error distribution is assumed, then the model is equivalent to the proportional-odds model (McCullagh & Nelder, 1989) and the within-level variances are fixed to $\pi^2/3$.

Measurement Model Formulations for Multilevel Covariance Structures. Goldstein and McDonald (1988) were the first to suggest estimating a common factor model at each data level:

$$\Sigma_B = \Lambda_B \Phi_B \Lambda'_B + \Theta_B$$
$$\Sigma_W = \Lambda_W \Phi_W \Lambda'_W + \Theta_W$$

(2)

Where the 'B' subscript refers to the between (school) level and the 'W' subscript denotes the within (student) level. Σ represents the model-implied population covariance matrix at a

given level. Each level has a unique factor loading matrix (Λ), factor intercorrelation matrix (Φ), and residual matrix (Θ). The within and between covariance models combine to form the total covariance structure for the total model:

$$\Sigma_T = \Lambda_B \Phi_B \Lambda'_B + \Lambda_W \Phi_W \Lambda'_W + \Theta_B + \Theta_W \qquad (3)$$

The model given in this equation shall be referred to as the Between and Within measurement model formulation.

The model simplifies greatly if it can be assumed that the factor structure at both levels is consistent. This assumption is testable. If the factor loading matrices at each level are constrained to equivalence such that $\Lambda_W = \Lambda_B = \Lambda$, the model equations simplify, and the interpretation of the latent variables is, appealingly, the same for both levels. In such a model, the construct is viewed as a single factor with variance that is parsed into school-level variation and student-level variation (Bollen, Bauer, Christ, & Edwards, in press; Goldstein & McDonald, 1988; Skrondal & Rabe-Hesketh, 2004). This restriction would be appropriate only if the factor structure at the within and between levels was equivalent.

The model is further simplified if there is no systematic unique variance in an item response that resides at the between level (i.e., the expected vale of Θ_B is zero). Imposing this restriction implies that, at the school level, the only true variance in an item results from the common factor, and not from any other source that is consistent for all students nested within a school. The major motivation for imposing this restriction is model parsimony. The factor indicators in this study should have specific individual-level variance; however, it would be difficult to conceive of a specific factor influencing the variables at the school-level.

22

If both of these restrictions are tested and are found to be permissible, the measurement model reduces to the more interpretable Variance Components formulation described by Rabe-Hesketh, Skrondal, and Pickles (2004), and Skrondal & Rabe-Hesketh (2004):

$$\Sigma_B = \Lambda\Phi_B\Lambda'$$
$$\Sigma_W = \Lambda\Phi_W\Lambda'+\Theta_W \qquad (4)$$
$$\Sigma_T = \Lambda\Phi_W\Lambda'+\Lambda\Phi_B\Lambda'+\Theta_W$$

Figures 1 through 4 present a comparison of the Between and Within formulation and the Variance Components formulation of measurement models for the latent constructs in this analysis. The Variance Components measurement model is nested within the Between and Within formulation but it requires that a parameter be fixed to a boundary value (Θ_B=0) such that regularity conditions for a likelihood ratio test are violated. Because the between-level residual variances are likely to be zero or very close to zero, this violation cannot be ignored. However, the models can be compared using fit statistics such as the AIC and the BIC.

Random Intercept Structural Models for Multilevel Data. In the structural model, student-level latent variables are regressed on other student-level latent variables, student-level measured variables, and school-level measured variables:

$$\eta_{ij} = B_W\eta_{ij} + \Gamma_W x_{ij} + \Gamma_B z_j + \xi_{ij} \qquad (5)$$

where η is random over students and schools and represents a latent variable, **B** contains the fixed regression coefficient(s) relating the latent variables, *x* represents a vector of observed variables measured at the student level, *z* represents a vector of observed variables measured at the school level, Γ_W contains fixed regression coefficients relating the student-level

measured variables to the endogenous latent variables, Γ_B contains fixed regression coefficients relating school-level measured variables to the endogenous latent variables, and ξ_{ij} is the unexplained factor variance.

The corresponding data model for the underlying true score of a given item (represented as y^*) is:

$$y_{ij}^* = \alpha_y + \Lambda\eta_{ij} + B\eta_{ij} + \Gamma_W x_{ij} + \Gamma_B z_j + \varepsilon_{ij} \qquad (6)$$

Where α_y represents the randomly distributed school mean (i.e., the random intercept) of that item (Bollen, et al., in press).

Comparing Models. If data are categorical and not MCAR, only the AIC and BIC (Akaike's information criterion and Bayesian information criterion) are currently available for evaluating the fit of disaggregated models with full-information maximum likelihood. When numerical integration is used, there is no covariance matrix available from which to derive a saturated model that is required for estimating a model chi square statistic (Asparouhov & Muthén, 2007; Mehta & Neale, 2005). A number of global fit indices rely on the chi-square, including the TLI, CFI, and RMSEA. Nested models can be compared with model chi-squares if a Satorra-Bentler scaling correction for nonnormality is used (Satorra & Bentler, 1999).

As an alternative to full information maximum-likelihood, the WLSMV estimator is available for multilevel categorical data in M*plus.* The WLSMV estimator uses a two-stage limited information approach to estimating model parameters as well as an unrestricted covariance matrix that can be used for obtaining global model fit indices that are chi-square based. However, this estimator is less efficient than the robust maximum-likelihood (MLR) alternative and requires missing data assumptions that have been shown to be inappropriate

for this data set (MCAR; Asparouhov & Muthén, 2007). Because global fit indices for MSEMs with categorical data are only currently available for the WLSMV estimator, fit indices will be reported for some models based on WLSMV solutions; however, parameter estimates, information criteria, and LRTs for MLR estimates will be reported because these estimates are more efficient and should be less biased under an MAR missing data mechanism.

Both the AIC and the BIC penalize for overparameterization in favor of more parsimonious models. The primary reason for using the BIC is to choose the "true" model for the data (i.e., given the data, which model has the highest likelihood). The AIC, on the other hand, is not based on the existence of a true model. Instead, the index maximizes predictive ability for future studies (Kuha, 2004). Both pieces of information provide unique and valuable information. Optimally, the best model would be both the most likely given the observed data and would have the greatest predictive validity.

Analysis Plan. Following the strategy of Muthén (1994), a relatively simple aggregated analysis was carried out to ensure that the hypothesized measurement and structural models fit reasonably well in the population. An aggregated analysis does not distinguish person-level and school-level variance, but it provides corrects standard error estimates for dependence within clusters by using a sandwich estimator (Muthén & Satorra, 1995). Aggregated analyses are useful in the preliminary stage of two-level covariance structure analysis because they are less computationally intensive and provide information for evaluating whether hypothesized measurement and structural models are reasonable.

The first step in the disaggregated analysis was testing the Variance Components measurement model formulations against the more complex alternative, the Between and

Within formulation. Once the Variance Components measurement models were determined to provide a superior fit (see Results section for details), the hypothesized model (Figure 5) was tested against an alternative model (Figure 6).

The hypothesized model draws upon a collection of theoretical models of the effects of diversity ranging from Allport's (1954) contact theory, the 2003 Supreme Court ruling on affirmative action, to a variety of recent empirical work. The model posits that racial diversity within a law school is directly and positively related to both individual intergroup contact and to diversity of ideas. In turn, both intergroup contact and diversity of ideas are hypothesized to lead to decreased perceptions of a meritocracy. These relationships should be present even after controlling for students' initial scores on perceptions of a meritocracy, their intergroup contact during college, age, gender, racial/ethnic minority status (White vs. non-White), political orientation, childhood relative income, the sector of the law school they attend, and the selectivity of the law school they attend.

Intergroup contact and diversity of ideas are hypothesized to be mediating variables in the model; however, they consist of retrospective measures assessed concurrently with the items measuring perceptions of a meritocracy at follow-up. This is a hindrance to the ability to infer causality because of the issue of selection (Holland, 1988). Despite the inclusion of all covariates that are likely to confound the relationship between racial diversity, intergroup contact, diversity of ideas, and perceptions of a meritocracy, it is possible that some unforeseen confounding variables may not be accounted for in the model. For this reason, an alternative model was tested that allowed correlations, rather than directional arrows, between all of the measures assessed at the second time point. Correlations were specified for both the within and the between parts of the correlational model so that the alternative

models are not chi-square equivalent. The model with a lower AIC and BIC provides a better fit to the data and is considered to be a more plausible representation of reality.

Multiple Groups Analysis. On average, 69% of the student bodies in the schools in both the core and volunteer samples are White. It is conceivable that the pattern of results might be moderated by the race/ethnicity of the student. While there were 1,113 White respondents at follow up, only 125 African American, 108 Asian American, 71 Mexican/Hispanic, and 116 multiracial students responded to the follow-up survey. Unfortunately, these numbers indicate that the statistical power to detect differences between non-White groups in a multiple groups analysis would be too low; however, given that Whites are generally the majority racial group in a school it is sensible to make a distinguish between Whites and non-Whites (e.g., ethnic/racial minorities).

As a first step in the multiple group multilevel SEM analysis, measurement invariance was tested for Whites and non-Whites following the procedure specified Millsap and Tein (2004) for categorical data. Measurement invariance for ordinal data requires that two criteria be met: 1) items must have the same relationship to the latent variance for both groups (i.e., equivalent factor loadings for all items; equivalently, no differential item functioning in the slope parameters), and 2) conditional on a particular value of the latent variable, Whites and non-Whites must endorse a level of the item at equal rates (i.e., equivalent thresholds for all levels on all items; no differential item functioning in the difficulty parameters). Once measurement invariance was verified (see Results section), two structural models were tested. The first constrained the structural paths of interest to equality (i.e., the paths from intergroup contact were constrained). The second model allowed these paths to be freely estimated across ethnic/racial groups. School-level effects are, by

definition, constant for all individuals within a school and therefore cannot vary across

ethnic/racial groups within schools. Model fit was compared using likelihood ratio tests.

Estimation. M*plus* v. 5 (Muthén & Muthén, 1998-2007) was used to estimate all

SEMs in this study. A full-information maximum-likelihood estimator for non-normal and

dependent data (MLR) was used for all analyses and the WLSMV estimator was used to

obtain global fit statistics as previously discussed. The MLR estimator is asymptotically

equivalent to the estimator proposed by Yuan and Bentler (2000). Gauss-Hermite quadrature

with five integration points was used to numerically evaluate the likelihood. The

combination of few items per each latent variable and many dimensions to integrate over

made the use of relatively few quadrature points a realistic, if less than desirable, option.

As of version 5, estimating the multiple groups model in M*plus* cannot be carried

out using the standard "grouping" at the same time that the "two level" command is used. As

an alternative, M*plus* estimates the equivalent of a multiple groups analysis as a finite

mixture model with complete training data (i.e., indicators of group membership).

Chapter 3

Results

This section briefly presents results from the aggregated analysis first, followed by results from several disaggregated models.

Aggregated Model (Table 4). The global fit (obtained using the WLSMV estimator) of the hypothesized diversity model indicated that the model-implied covariance structure was a close approximation to the population covariance structure. The null hypothesis of perfect model fit was rejected ($\chi^2_{(25)}$=155.77). This finding is not surprising given the large sample size: all models are approximations to the population (e.g., MacCallum, Browne & Cai, 2007). However, the estimated CFI and TLI were well above .90 (.95 and .94, respectively), and the estimated RMSEA was .05. Interestingly, the choice of estimator (MLR vs. WLSMV) made little difference in terms of standardized structural estimates, significance levels, or direction of effects in the aggregated model. Parameter estimates obtained with the MLR estimator are reported.

Factor loadings ranged from 1.00 to 1.81 on the perceptions of a meritocracy factor and from 1.00 to 2.28 on the diversity of ideas factor. Structural parameter estimates for the population-average model indicate that racial diversity (RDI) significantly increases contact between students of different race/ethnicities (β=3.12; *SE*=.60; std β = .24). Standardized regression coefficients (labeled std β) are reported to provide an indication of the effect size. These coefficients are equivalent to standardized regression coefficients in typical regression

analysis. RDI does not significantly predict diversity of ideas or perceptions of a meritocracy in the aggregated model. Intergroup contact was a hypothesized mediator of the effect of RDI on perceptions of a meritocracy and on diversity of ideas. This hypothesis was supported in the preliminary model: Intergroup contact significantly predicted reduced perceptions of a meritocracy (β=-.06; *SE*=.02; std β = -.07) and diversity of ideas (β=.11; *SE*=.02; std β = .17). Recall that these effects control for intergroup contact during college, pre-test perceptions of a meritocracy, race, gender, age, childhood SES, and school selectivity. Neither LSAT scores nor school sector were significantly related to the endogenous variables and so they were eliminated from the model. It was also hypothesized that diversity of ideas would be negatively causally linked to perceptions of a meritocracy at follow-up; however, this effect was nonsignificant.

The model described above provides some valuable information. For instance, it gives a rough idea of which covariates might be necessary to include as predictors of various endogenous variables in the disaggregated model. However, there are serious drawbacks to this model. Most importantly, student-level effects are confounded with school-level effects (e.g., Muthén & Satorra, 1995; Raudenbush & Bryk, 2002). Every variable measured at the person-level has some proportion of variance that is constant for all individuals clustered within a school (i.e., school-level variance). While school-level predictors only explain variance at the between level, predictors at the individual level predict some proportion of between variance and some proportion of within variance. Under the aggregated approach, it is not known whether the factor structure at the within-level and between-level is consistent, or whether structural paths are consistent across levels. Disaggregating the variance between

and within provides a more precise understanding of the specific person- and school-level mechanisms underlying the effects of educational diversity in law schools.

Disaggregated Measurement Models. The first step in conducting a disaggregated analysis was to determine the correct measurement models for the latent variables. Recall that there was a choice between the Between and Within formulation and the Variance Components formulation (Bollen, et al., in press). The Variance Components formulation is more parsimonious and is more intuitive than the Between and Within formulation but makes some strong, albeit testable, assumptions. Namely, the Variance Components model assumes that: (1) all of the unique variance in measured variables resides at the person-level (i.e., there is no specific school-level variance in an indicator that does not result from the common factor); and (2) the common factor loading structure is invariant across levels of nesting. From a practical standpoint, the Between and Within measurement models require a minimum of several hours each to converge while the Variance Components models could be estimated in a matter of seconds. Computing time drastically increases as more random effects are introduced into a structural model. Also, the Variance Components estimates were more stable when using numerical integration, as required here by the ordinal nature of the indicators.

For the diversity of ideas construct, two between-level residual variances in the Between and Within model had to be fixed to zero to avoid Heywood cases (boundary variance estimates). This adjustment alone provides some evidence that the Variance Components model is more appropriate because one requirement of this formulation is that between-level residuals are zero. After these residual variances were fixed, standard error results from the MLR estimator could not be estimated because the solution was a potential

saddle point. Because of this, the M*plus* program calculated standard errors using a first-order Taylor series approximation. The information criteria were very slightly lower for the Between and Within formulation (AIC = 20,309.63, BIC = 20,446.06, sample-size adjusted BIC = 20,369.81) than for the Variance Components formulation (AIC = 20,381.23; BIC = 20,489.24; sample-size adjusted BIC = 20,428.88); however, in light of the Heywood cases and instability of the Between and Within solution, the Variance Components formulation was selected. Table 5 contains factor loadings and threshold estimates for the Variance Components formulation of the diversity of ideas measurement model.

Measurement model selection for the perceptions of a meritocracy factor was more straightforward. A single between-level residual had to be fixed to zero in the Between and Within formulation measurement model but there were otherwise no convergence issues. The information criteria indices indicated that the Variance Components model provided a superior fit (AIC = 16,020.54, BIC = 16,111.48, sample-size adjusted BIC = 16,060.64) to the Between and Within model (AIC = 16,023.50, BIC = 16,137.18, sample-size adjusted BIC = 16,073.64). Table 6 contains parameter estimates for the Variance Components formulation of the perceptions of a meritocracy measurement model.

The intra-class correlation (ICC; Muthén, 1994) is estimated for latent variables by dividing the estimated variance at the between-level into the total latent variable variance (between plus within) in the unconditional measurement model. As noted by Raudenbush, Roawn, and Kang (1991), the intra-class correlations are generally larger for latent variables than for measured variables due to the disattenuation for measurement error. The ICC for diversity of ideas is fairly low (.03). It is higher (.14) for the perceptions of a meritocracy factor. The ICC indicates the proportion of the variance in the latent variable that resides at

the level of the school. In other words, it is the correlation between students resulting simply from attending the same law school. The ICC for intergroup contact during law school was estimated to be .22 in SAS Proc Mixed.

Alternative Models. Two structural models were compared under the assumption of population homogeneity (i.e., a single group model for Whites and non-Whites). Model comparisons were necessary in this analysis because the non-experimental survey design does not allow researchers to draw firm conclusions about causality (Cook & Campbell, 1979). All plausible models should be tested and considered to rule out substantively different alternatives (MacCallum, Wegener, Uchino, & Fabrigar, 1993). Thus, a model with correlations between all constructs measured in the follow-up survey at both the within and the between levels was considered as an appropriate alternative to the model with directional paths from intergroup contact to diversity of ideas at the within level, intergroup contact to perceptions of a meritocracy at the within level, and from diversity of ideas to perceptions of a meritocracy at the within level. The alternative models considered here are not nested and so cannot be compared using a LRT. BIC and AIC indices are available for the MLR-estimated models and were designed for the purpose of comparing fit of non-nested models. Global fit indices from WLSMV-estimated models were also used as a heuristic for comparing model fit.

Parameter estimates for the correlational model (depicted in Figure 6) are presented in Table 7. Many of the structural paths originating from exogenous covariates were fixed in this and subsequently reported models because there were not enough independent sampling units in the sample to derive stable estimates of all of the parameters in the model if these paths were freely estimated. Values were fixed to parameter estimates from single-factor

33

Multiple Indicator Multiple Cause (MIMIC; Bollen, 1989) models. All paths that involved perceptions of a meritocracy, diversity of ideas, and intergroup contact were estimated in all models. Some paths from *RDI* were fixed to zero on the basis of MIMIC models as necessitated for convergence of the between-level covariance matrix. The alternative model fits well. On 29 degrees of freedom, the chi-square distributed discrepancy function for the alternative model is 124.69 with a CFI and TLI of .96 and an RMSEA of .04.

The nonsignificant correlation between perceptions of a meritocracy and diversity of ideas at the within level indicated that the hypothesized directional path between diversity of ideas and perceptions of a meritocracy is not supported (see Table 7). Thus, it was not estimated in the test of the hypothesized model. The (revised) hypothesized model (full results shown in Table 8) was more generalizable and was more likely given the data than the alternative correlational model. The AIC for the correlation model was 40,890.12 and was 40,819.39 for the hypothesized directional model. The BIC for the correlation model was 41,024.79 and was 40,949.17 for the directional model. Recall that lower information criterion values reflect a better fit relative model fit. Also on 29 degrees of freedom, the WLSMV-estimated chi-square distributed discrepancy function for the revised hypothesized model was 118.17 with a CFI of .97, a TLI of .96, and an RMSEA of .04. While this model is not formally nested with the alternative model, it is clear from the global fit statistics that the hypothesized model performs slightly better than its competing alternative. Thus, parameter estimates from this "best fitting" (single group) model will be described in detail in the text. As was the case with the aggregated model, standardized parameter estimates are quite similar when estimated with the WLSMV estimator and with the MLR estimator.

Best Fitting Model. Racial diversity (*RDI*) was expected to directly affect intergroup contact during law school, diversity of ideas, and perceptions of a meritocracy at follow-up. Indeed, RDI significantly increased intergroup contact during law school (β = 3.20; *SE* = .66; std β = .61) and was a marginally significant direct predictor of diversity of ideas (β = .46; *SE* = .28; std β = .24). The direct effect of RDI on perceptions of a meritocracy at follow-up was fairly strongly negative (β = -.60; *SE* = .16; std β = -.50). Racial diversity also indirectly affected diversity of ideas and perceptions of a meritocracy through its relationship with intergroup contact during law school. The effect of intergroup contact on diversity of ideas is small to moderate, but significant and positive (β=.17; SE=.02; std β = .20), and its effect on perceptions of a meritocracy is small but significantly negative (β=-.10; SE=.01; std β = -.14). Inferential tests of indirect effects in two-level models are not currently available.

Perceptions of a meritocracy was negatively correlated with diversity of ideas at the school level (β=-.04; SE=.02; std β = -.90) but these constructs were not significantly correlated at the within level. Directional effects of between variables measured at the individual-level are not estimable at the between level as a limitation of the estimation method currently used in M*plus* (version 5). The parameterization of multilevel SEMs does not incorporate within-level variables in the between-level predictor-side matrix. To allow such a model would require a more complex restructuring of the model.

The proportion of between variance in intergroup contact during law school grew to .39 in the final model, a large increase from the unconditional ICC of .22. This change is an indication that more of the within-level variation than between-level variation in intergroup contact has been explained by the covariates in the final model. The proportion of variance at the between-level in the final model for perceptions of a meritocracy is .65, a large increase

from .14 in the unconditional model. The proportion of variance at the between level had a very slight increase from .03 in the unconditional model to .04 in the final model. These statistics indicate that demographics (age, gender, race/ethnicity, political orientation, and childhood relative income), as well as baseline perceptions of a meritocracy and intergroup contact during college, explain a larger proportion of the variance in intergroup contact during law school, perceptions of a meritocracy at follow-up, and diversity of ideas than school-level predictors (RDI and selectivity). Intergroup contact during law school also predicts some of the individual-level variation in the endogenous latent variables.

Measurement Invariance across Race/Ethnicity. Once the causal model was selected, racial group invariance was tested. Measurement invariance is a prerequisite for making inferences about structural invariance (Millsap, R.E & Meredith, W., 2007). A likelihood ratio test (LRT) with the Satorra-Bentler scaling correction factor for nonnormality was nonsignificant for both diversity of ideas ($\chi^2_{(13)}$=1.07) and for perceptions of a meritocracy ($\chi^2_{(9)}$=1.13). Two structural models were compared: one with freely estimated paths for all within-level parameters and one with important structural paths constrained to equality across groups. The corrected LRT was nonsignificant ($\chi^2_{(3)}$=1.52) indicating structural invariance of the mechanism of diversity for Whites and non-Whites. Thus, the "best fitting" single group model remains the most parsimonious model of the mechanism of how racial diversity influences student experiences with diversity in law school, and how attitudes about socio-political issues are influenced by racial heterogeneity within a school, as well as experiences with diversity.

Chapter 4

Discussion

In this manuscript, I aimed to simultaneously contribute to a substantively important question by employing the most appropriate statistical methodology that is currently available. In doing so, I put the developing methodology to the rigorous test or real-world data analysis. As such, there are two main components to the discussion. The first section concerns findings related to educational diversity, and the second section discusses my experience with implementing MSEM. Practical advice for applied researchers is given, along with suggested directions for future developments in the methodology.

Educational Diversity

Support for Hypotheses. It is important to note that although the link between school-level racial diversity and student-level contact may seem trivial, it should not be taken for granted. Whitehead and Wittig (2004) found that some students resist multiculturalism in schools by choosing to self-segregate. Springer (1996) called Allport's (1954) Contact Theory into question based on findings that college students who are not receptive to diverse interactions self-segregate when faced with racially diverse environments. The significant causal link between school-level racial diversity and intergroup contact during law school was significant even after controlling for contact during college and for other background factors. Thus, this study provides evidence to suggest that self-segregation does not completely hinder the positive effects of racial diversity in law schools. As an additional

note, the possibility of selection bias of students into racially diverse schools was tested in the EDP sample. Using LSAC data on admissions decisions, we determined that prospective students were driven primarily by school selectivity and that racial diversity was uncorrelated with students' decisions to matriculate to a particular law school.

This study was conducted to test several hypotheses about educational diversity in law schools. First, on the basis of Allport's (1954) contact theory of prejudice reduction, I believed that higher levels of intergroup contact during the law school years would lead to decreased prejudices about disadvantaged populations. This hypothesis was fully supported by the data. Controlling for students' predispositions to intergroup contact with diverse peers, baseline prejudices, political orientation, ethnicity/race, gender, and childhood SES, students who interacted with a diverse group of peers during law school had less prejudiced beliefs at the completion of law school. In addition, intergroup contact partially mediated the effect that school-level racial diversity had on reducing prejudice.

Allport (1954) theorized that the presence of diverse groups is necessary but not sufficient for reduction in prejudice against other groups. Rather, he said that interactions should be cooperative, supported by an authority, and that all parties should be of equal social status. Law students are generally of equivalent social standing, and if they choose to interact during law school, their interactions are more likely to be cooperative (e.g., studying and socializing together) than competitive. Little is presently known about how supportive law school faculty and administrators in these law schools were; however, research by Pettigrew and Tropp (2006) suggests that gains can still be made without this support. The effectiveness of administrative support for intergroup contact is an important topic for future study.

The effect of intergroup contact during law school on perceptions of a meritocracy at follow-up was small but statistically significant. An increase in one standard deviation of intergroup contact during law school is associated with a decrease in .10 standard deviations on the perceptions of meritocracy scale at follow-up. The recent meta-analysis by Pettigrew and Tropp (2006) suggested a moderate effect size. The effect in the law school sample may not have been as large for a few reasons. First, law school is a competitive environment and not all student interactions may be positive. Second, we do not have information on how much these interactions were supported by law school faculty and administrators. Third, the average student entering law school is in his or her mid-to-late twenties with more crystallized belief systems than those of younger students, making law student attitudes and beliefs more difficult to influence.

The second hypothesis was that schools with more racial diversity would be characterized by a more robust exchange of diverse ideas. There was a moderate but nonsignificant direct effect of racial diversity on diversity of ideas, as well as a substantial indirect effect of racial diversity on diversity of ideas via intergroup contact. The statistical significance of indirect effects cannot be tested across levels of the MSEM; however, both paths in the indirect effect are moderate-to-large and are statistically significant. These effects persisted after controlling for student demographics and school characteristics (e.g., selectivity, sector).

This finding suggests that racial diversity at a school is not in and of itself sufficient to promote a diverse exchange of ideas, but the extent to which there are intra-personal interactions across racial/ethnic groups in a school, a robust exchange of ideas follows. This effect is not possible without the presence of school-level racial diversity.

Theory is mixed regarding the relationship between the diversity of ideas to which students are exposed to and their prejudiced attitudes. For instance, Chang (2002) found that students who were compulsorily exposed to courses dealing with issues of multiculturalism had reductions in racism compared to students who did not have such exposure. On the other hand, Federico and Sidanius (2002) and Reyna, et al. (2005) presented empirical support for the notion that prejudices can be strengthened and supported with exposure to higher education. It is perhaps not surprising, then, that I found no significant relationship between diversity of ideas and perceptions of a meritocracy at the individual level. There was a significant negative correlation between the constructs at the school-level, indicating that schools that are characterized by a more diverse exchange of ideas have lower average levels of prejudice. The standardized effect of this correlation was quite large (.90); however, directional hypotheses cannot be tested at the school-level in current software and both directional effects are plausible. All that can be inferred from this finding is that low levels of prejudice in a school coexist with a more robust exchange of ideas within a school.

Policy Implications. This study was motivated by the recent challenges to the use of affirmative action in education. There were two purposes: 1) to provide empirical support for the benefits of racial diversity in education, and 2) to determine mediating mechanisms that have the potential to be manipulated by school administrators. The latter, while beneficial for any academic administrator wishing to boost the effects of educational diversity, will become particularly important should race-conscious admissions policies be completely banned.

The first major substantive implication of this study is that racial diversity provides *measurable* benefits for students. Furthermore, the finding of no ethnic/racial group differences in the structure of the effects suggests that these benefits accrue for students of all

40

race/ethnicities. Assuming that the mission of the majority of institutions of higher education is to facilitate the exchange of diverse ideas, administrator should focus on achieving a racially diverse student body to the extent possible. Increasing the racial heterogeneity of the student body will also disabuse law students of prejudiced attitudes before they enter into the workforce.

Second, the finding that the benefits of racial diversity are partially mediated by intergroup contact suggests a mechanism through which educators can maximize the benefits of racial diversity within their school while working with a fixed amount of racial diversity in the student body. By actively encouraging cooperative interactions between students of different ethnic and racial backgrounds, students will be exposed to a wider array of new ideas and out-group prejudices will dissipate. For example, law schools may encourage students to form discussion groups with peers whose viewpoints differ from their own. This particular technique is supposed by an experimental study conducted by Antonio, et al. (2004), who found that students that were randomly assigned to racially diverse focus groups developed a higher degree of integrative complexity in their perspectives on an array of issues compared to the racially homogenous groups. Administrators may also encourage student group leaders to actively recruit students of diverse racial/ethnic backgrounds to participate in extracurricular student groups (e.g., mock trial, student journals).

Limitations. It is not legally possible to manipulate experimentally racial diversity in law schools so causality cannot be proved. This limitation was minimized with the inclusion of potentially confounding covariates, a longitudinal design, and with alternative model comparisons. Furthermore, self-selection of students into racially diverse schools was

eliminated as a possibility by examination of admissions and matriculation records. This finding is consistent for both genders and all racial/ethnic groups in the study.

Current State of Methodology.

For the most part, the ability to use MSEMs to model data offers a tremendous advantage over other commonly used approaches to modeling structural and nested processes (e.g., multilevel modeling without latent variables, structural equation modeling without accounting for nesting, or that account for nesting by using a population-average aggregated approach). Indeed, few diversity researchers even use these approaches, instead opting to conduct multiple regressions that do not control for nesting (e.g., Antonio, 2001; Gurin, et al., 2002), or correlation coefficients (e.g., Chang, 2002).

Multilevel modelers are very familiar with the consequences of ignoring nestedness in data. Researchers who model multilevel data as single level data are plagued with inflated standard errors and high rates of type I error (e.g., Bryk & Raudenbush, 2002). Structural equation modelers are equally well versed in the pitfalls inherent in failing to use common factor models to model unobserved constructs. Using measured variables, including composites, in the place of latent constructs results in lower reliability and reduced construct validity (e.g., Hoyle & Robinson, 2004). Measurement models, on the other hand, provide a statistical way to implement multitrait-multimethod measurement (Campbell & Fiske, 1959) and to model directly unique error variance in items (Thurstone, 1947). Psychologists frequently encounter nested data, either due to sampling strategies or because of naturally occurring clusters in the population (such as students nested in schools). Often, the proportion of variation in a variable that is attributable to each level of nesting provides interesting information. Just as often, psychologists are interested in measuring constructs

that are not directly observable. MSEM, to the extent that it is theoretically developed and computationally practical, is able to handle both of these situations simultaneously.

There is a major advantage to using a disaggregated (i.e., multilevel) SEM approach to handling complex survey data over the aggregated (i.e., population average) approach. The population-average approach is useful for drawing inferences about average trends that may appear in a population but it is not useful for making inferences about a particular individual in the population (Muthén & Satorra, 1995). The disaggregated/multilevel approach enables more finely tuned inferences by parsing variance at the between- (school-) level from the variance at the within- (student-) level. This provides two pieces of information: (1) the correlation between students that is associated with attending the same school (intraclass correlation; ICC), and (2) the variability that exists across schools and across individuals around the population average.

MSEM is not new, but its incorporation into statistical software is relatively recent. As such, there are very few instances of applied uses of MSEM in the literature (exceptions being pedantic illustrations authored by developers of the approach). With the exclusion of Skrondal and Rabe-Hesketh's (2004) textbook, there are no published examples of conducting a MSEM with categorical data. Skrondal and Rabe-Hesketh (2004) did not mention issues of missing data.

Computational Issues. Numerical integration is required to implement maximum likelihood estimation when data are categorical. Item response theorists, who have conducted nonlinear factor analysis for decades, have developed a number of numerical estimation techniques that provide unbiased, precise, and stable estimates when only one or two random effects are present (see Swygert, McLeod, & Thissen, 2001; Wirth & Edwards, 2007).

Researchers dealing with only a couple of factors have a choice between using upwards of 15 quadrature points, or using fewer adaptive quadrature points. Multilevel structural equation modelers typically do not have such a luxury.

In this analysis, I was extremely frugal with allowing dimensions of integration. Instead of using a measurement model for baseline perceptions of a meritocracy, I computed a composite score. Instead of declaring intergroup contact as Poisson-distributed at baseline and at follow-up, I treated it as a continuous variable. The Variance Components measurement model formulation was used instead of the Between and Within measurement model formulation. Despite these choices, there were still five dimensions of numerical integration required to evaluate the likelihood of the structural models. Adaptive quadrature was not a viable option, and non-adaptive quadrature points had to be limited to avoid allowing the exponential growth in computing time to become unwieldy. Each additional quadrature point increases points of integration by an additive power, so five quadrature points were generally used for model estimation. Schillings and Bock (2005) were able to recover population parameter values with as few as two adaptive quadrature points, providing some credence to the idea that population parameter values can be recovered with five non-adaptive quadrature points.

A MSEM would be extremely difficult to implement if the Between and Within measurement model formulation were used. Each single factor measurement model took several days each to converge, and often resulted in problematic solutions (e.g., Heywood cases, saddle points). Whenever possible, the more interpretable Variance Components model should be used.

Limitations on Testing Theorized Model. I encountered two barriers to theory testing while using MSEM. First, only variables with no variation at level one can be used to predict between-level variation in other variables. For instance, only the within-level directional effect of diversity of ideas on perceptions of a meritocracy could be specified; covariation between these factors was allowed at the between level, but predictive paths were not permitted. The second limitation was relatively minor. The standard error of cross-level indirect effects is not estimable in M*plus*. Thus, the statistical significance of the indirect effects of racial diversity on perceptions of a meritocracy and on diversity of ideas was not tested.

Chapter 5

Conclusion

Educational Diversity. Educational institutions have a long way to go before ethnic

and racial groups receive equal representation (Isaacs, 2008). This study provides empirical

evidence that ethnic/racial diversity benefits students of all ethnic/racial backgrounds, over

and above the economic benefits that clearly exist for disadvantaged groups. Meanwhile, the

use of race conscious admissions decisions is becoming increasingly stringent (*Gratz vs.*

Bollinger, 2003; *Meredith vs. Jefferson Co. Board of Ed.*, 2007; *Parents vs. Seattle*, 2007).

Therefore, educational administrative policy would benefit from research that focuses on

enhancing educational diversity with race neutral policies. Shultz (2007) recommends

increasing racial diversity in schools by evaluating prospective students based on

"effectiveness factors." The current study recommends that administrators focus efforts on

increasing positive interactions between students of different backgrounds. Once

implemented, the effectiveness of these strategies should be evaluated.

Multilevel Structural Equation Modeling with Categorical and Missing Data.

Researchers now have at their disposal a useful tool for modeling the structural relationships

between latent variables that are measured with non-normally distributed variables that

contain variation at more than one level. MSEM is an important development for accurate

representation of theoretical processes. As previously discussed, there is some room for

improvement in software and technical documentation. Given the rate of new technological

developments, MSEM shows great promise as a technique to be regularly implemented in

applied research.

Table 1

Law School Characteristics for Volunteer and Core Baseline Samples

	Mean/Proportion		SD		Min		Max	
	Volunteer	Core	Volunteer	Core	Volunteer	Core	Volunteer	Core
RDI	.46	.43	.10	.16	.23	.17	.66	.71
Private	.50	.54						
Enrollment	923.56	717.78	556.88	340.07	361	220	2868	1667
Selectivity	.30	.27	.11	.08	.16	.12	.60	.47

Note. RDI = Racial Diversity Index, a measure of the racial heterogeneity (the compliment of the sum of the squared proportion of each ethnic group); unity indicates complete heterogeneity and zero indicates pure homogeneity. Selectivity is calculated as the number admitted over the number of applications so lower numbers indicate greater selectivity. Core sample N = 50 schools; Volunteer sample N = 16 schools.

Table 2

Demographic Characteristics of Baseline Participants in Volunteer and Core Samples

	Mean/Proportion		SD		Min		Max	
	Vol.	Core	Vol.	Core	Vol.	Core	Vol.	Core
Female	.52	.52						
Age	25.41	25.42	5.04	5.15	19	18	58	61
Childhood Household Income	3.49	3.37	1.52	1.39	1	1	8	8
LSAT	159.15	156.68	5.47	7.04	136	120	179	180
Political Orientation	2.56	2.69	.93	.97	1	1	5	5
Married or Civil Union	.17	.18						
Race/Ethnicity								
White	.69	.68						
African American	.06	.10						
Asian-American	.10	.09						
Mexican-American or Hispanic	.05	.05						
Multiethnic	.09	.08						

Note. Childhood household income and ranges from 1 (*Below $10,000*) to 8 (*Over $500,000*). Political Orientation ranges from 1 (*Extremely Liberal*) to 5 (*Extremely Conservative*). Volunteer sample consists of 1963 people. Core sample consists of 6,100 people.

Table 3

Attrition Analysis

Item	Attriters N=5,883	Completers N=2,180
Gender		
Female	50.3%	56.7%
Race/Ethnicity		
African American	9.5%	7.1%
Asian American	9.4%	7.1%
Mexican/Hispanic	5.3%	4.9%
Multiracial	8.6%	8.1%
White	67.1%	72.8%
Political Orientation		
Extreme Liberal	8.1%	11.9%
Liberal	36.0%	42.5%
Moderate	32.2%	26.3%
Conservative	18.7%	15.9%
Extreme	2.5%	2.1%
Conservative		
Relative Childhood Income		
Very Low	4.2%	4.1%
Low	12.8%	14.2%
Average	34.9%	35.0%
High	41.1%	41.5%
Very High	7.1%	5.1%
"Because Irish, Italians, Jews, and many other minorities overcame prejudice and worked their way up, Blacks should do the same without any special favors."		
Strongly Disagree	16.0%	18.6%
Disagree	30.2%	35.3%
Neutral	30.1%	26.7%
Agree	16.9%	14.4%
Strongly Agree	5.5%	4.6%

Note. Chi-square test statistics indicate that endorsement patterns differ significantly for attriters and completers on all of these items.

Table 4
Final Aggregated Model

Outcome	Predictor	Regression Weight (SE)
Perceptions of a meritocracy (Time 2)	**Racial Diversity Index**	-.44(.34)
	Intergroup contact during Law School	-.07(.02)***
	Perceptions of a meritocracy (Time 1)	.44(.02)***
	Intergroup contact in College	.0(-)[a]
	Female	-.38(.06)***
	White	.45(.09)***
	Age	.0(-)[a]
	Politically Conservative	.0(-)[a]
	Relative Childhood Income	.0(-)[a]
	Percent of Students Accepted	1.16(.48)*
Diversity of ideas	**Racial Diversity Index**	-.18(.28)
	Intergroup contact during Law School	.11(.02)***
	Perceptions of a meritocracy (Time 1)	-.05(.02)**
	Intergroup contact during College	.0(-)[a]
	Female	.0(-)[a]
	White	.30(.07)***
	Age	.01(.01)
	Politically Conservative	.0(-)[a]
	Relative Childhood Income	.08(.03)*
	Percent of Students Accepted	.0(-)[a]
Intergroup contact during Law School	**Racial Diversity Index**	3.12(.60)***
	Perceptions of a meritocracy (Time 1)	-.05(.02)**
	Intergroup contact during College	.18(.02)***
	Female	.0(-)[a]
	White	.0(-)[a]
	Age	.03(.01)***
	Politically Conservative	.07(.05)
	Relative Childhood Income	-.20(.04)***
	Percent of Students Accepted	.0(-)[a]
Perceptions of a meritocracy (Time 2) with Diversity of ideas		-.05(.05)

Note. Bolded predictors indicate hypothesized paths; AIC=40,797.62; BIC=41,101.40; n-adjusted BIC=40,929.84 from MLR estimator.
$\chi^2_{(24)}$=98.08; CFI=.97; TLI=.97; RMSEA=.04 from WLSMV estimator.
[a] Fixed to zero based on earlier models. *p<.05; **p<.01; ***p<.001

Table 5

Diversity of ideas Variance Component Measurement Model

	Loading (SE)
Quality of Class Discussions	1.00(-)
Respectful Exchange of Political Views	1.58(.12)***
Expression of Diverse Belief	2.34(.24)***
School is Open to New Ideas	1.09(.08)***
Thresholds	
Discuss 1	-3.07
Discuss 2	-.31
Discuss 3	3.00
Respect 1	-5.17
Respect 2	-2.71
Respect 3	-.07
Respect 4	3.08
Belief 1	-4.41
Belief 2	-1.18
Belief 3	2.95
Open 1	-3.94
Open 2	-2.10
Open 3	-.15
Open 4	2.26

Note. AIC=20,381.23; BIC=20,489.24; n-adjusted BIC=20,428.88. ICC=.03

Table 6

Perceptions of a meritocracy Variance Component Measurement Model

	Loading (SE)
In America today, everyone has an equal opportunity to succeed	1.00(-)
Because Irish, Italians, and Jews worked their way up, Blacks should do the same without any special f avors	2.14(.18)***
People on the bottom of the socio-economic scale are probably lazier than people at the top	1.81(.14)***
Thresholds	
Equal 1	-3.27
Equal 2	.12
Equal 3	3.10
Equal 4	5.54
Favor 1	-.84
Favor 2	1.32
Favor 3	2.82
Favor 4	5.14
Lazy 1	-2.49
Lazy 2	1.35
Lazy 3	2.86
Lazy 4	5.65

Note. AIC=16,020.54; BIC=16,111.48; n-adjusted BIC=16,060.64. ICC=.14

Table 7

Single Group Disaggregated Two-Level Model with Correlations between Time 2 Outcomes

Outcome	Predictor	Regression Weight (SE)
Perceptions of a meritocracy (Time 2)	**Racial Diversity Index**	-2.41(.13)***
	Perceptions of a meritocracy (Time 1)	.25(.01)***
	Intergroup contact in College	.0(-)[a]
	Female	-.28(.04)***
	White	.27(.06)***
	Age	.0(-)[a]
	Politically Conservative	.21(.03)***
	Relative Childhood Income	.0(-)[a]
	Percent of Students Accepted	.0(-)[a]
Diversity of ideas	**Racial Diversity Index**	3.17(.36)***
	Perceptions of a meritocracy (Time 1)	-.04(.02)*
	Intergroup contact during College	.0(-)[a]
	Female	.0(-)[a]
	White	.31(.07)***
	Age	.02(.00)***
	Politically Conservative	.0(-)[a]
	Relative Childhood Income	.08(.03)*
	Percent of Students Accepted	2.44(.40)***
Intergroup contact during Law School	**Racial Diversity Index**	4.50(.90)***
	Perceptions of a meritocracy (Time 1)	-.01(.02)
	Intergroup contact during College	.18(.02)***
	Female	.0(-)[a]
	White	-.19(.09)*
	Age	.03(.01)***
	Politically Conservative	.0(-)[a]
	Relative Childhood Income	-.13(.04)***
	Percent of Students Accepted	.13(1.17)
Perceptions of a meritocracy with Diversity of ideas Within		-.04(.04)
Perceptions of a meritocracy with Diversity of ideas Between		-.14(.04)***
Perceptions of a meritocracy w/ Intergroup contact Within		-.15(.04)***
Perceptions of a meritocracy w/ Intergroup contact Between		-.06(.05)
Diversity of ideas with Intergroup contact Within		.37(.06)***
Diversity of ideas with Intergroup contact Between		.08(.07)

Note. Bolded predictors indicate hypothesized paths; AIC=40,890.12; BIC=41,199.53; n-adjusted BIC=41,024.79 from MLR estimator. [a]Fixed to zero for identification purposes—nonsignificant in previous analyses. $\chi^2_{(29)}$=124.69; CFI=.96; TLI=.96; RMSEA=.04 from WLSMV estimator; *p<.05; **p<.01; ***p<.001

Table 8

Single Group Disaggregated Two-Level Model with Causal Paths between Time 2 Outcomes[b]

Outcome	Predictor	Regression Weight (SE)
Perceptions of a meritocracy (Time 2)	**Racial Diversity Index**	-.60(.16)***
	Intergroup contact (Law)	-.10(.01)***
	Perceptions of a meritocracy (Time 1)	.26(.01)***
	Intergroup contact (College)	.0(-)[a]
	Female	-.26(.04)***
	White	.28(.06)***
	Age	.0(-)[a]
	Politically Conservative	.23(.03)***
	Relative Childhood Income	.0(-)[a]
	Percent of Students Accepted	.0(-)[a]
Diversity of ideas	**Racial Diversity Index**	.46(.28)†
	Intergroup contact (Law)	.17(.02)***
	Perceptions of a meritocracy (Time 1)	-.04(.02)*
	Intergroup contact (College)	.0(-)[a]
	Female	.0(-)[a]
	White	.32(.07)***
	Age	.02(.01)**
	Politically Conservative	.0(-)[a]
	Relative Childhood Income	.09(.04)*
	Percent of Students Accepted	1.64(.37)***
Intergroup contact during Law School	**Racial Diversity Index**	3.20(.66)***
	Perceptions of a meritocracy (Time 1)	-.02(.02)
	Intergroup contact (College)	.19(.02)***
	Female	.0(-)[a]
	White	-.40(.11)***
	Age	.03(.01)***
	Politically Conservative	.0(-)[a]
	Relative Childhood Income	-.14(.04)***
	Percent of Students Accepted	.0(-)[a]
Perceptions of a meritocracy w/ Diversity of ideas Within		-.02(.04)
Perceptions of a meritocracy w/ Diversity of ideas Between		-.04(.02)**

Note. Bolded predictors indicate hypothesized paths; AIC=40,819.39; BIC=41,117.55; n-adjusted BIC=40,949.17 from MLR estimator. [a]Fixed to zero for identification purposes—nonsignificant in previous analyses. [b]No causal path between Diversity of ideas and Perceptions of a meritocracy was estimated given the non-significant correlation in correlational model.. $\chi^2_{(29)}$=118.17; CFI=.97; TLI=.96; RMSEA=.04 from WLSMV estimator; *p<.05; **p<.01; ***p<.001
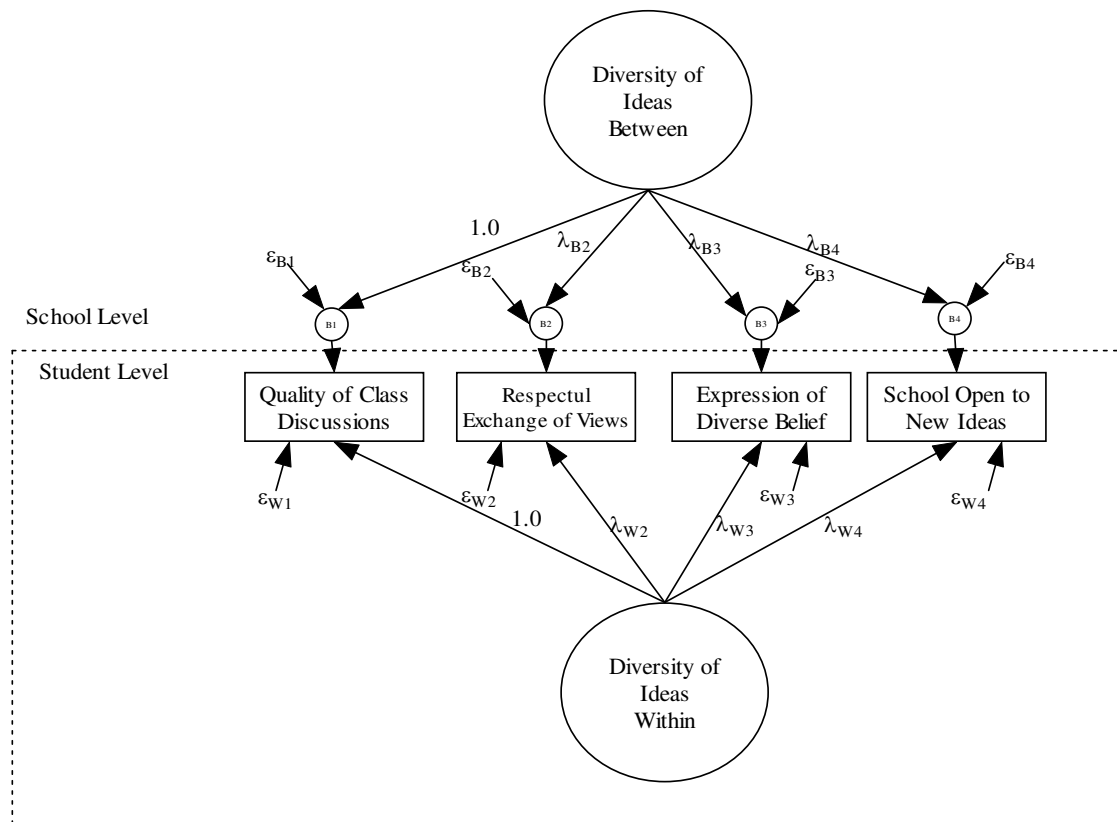
Figure 1. Between and Within measurement model formulation for diversity of ideas

construct. Individual-level item variances are implicitly fixed to $\pi^2/3$ by the logit function

relating the ordinally-measured items to the factor. Between-level item intercept variances
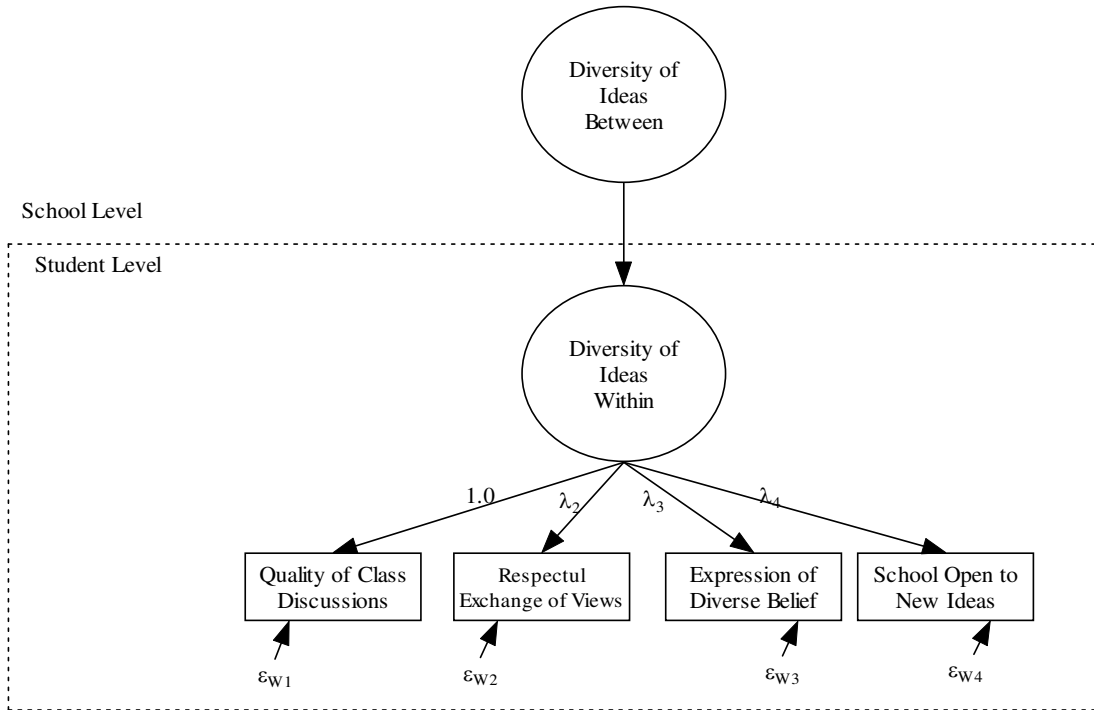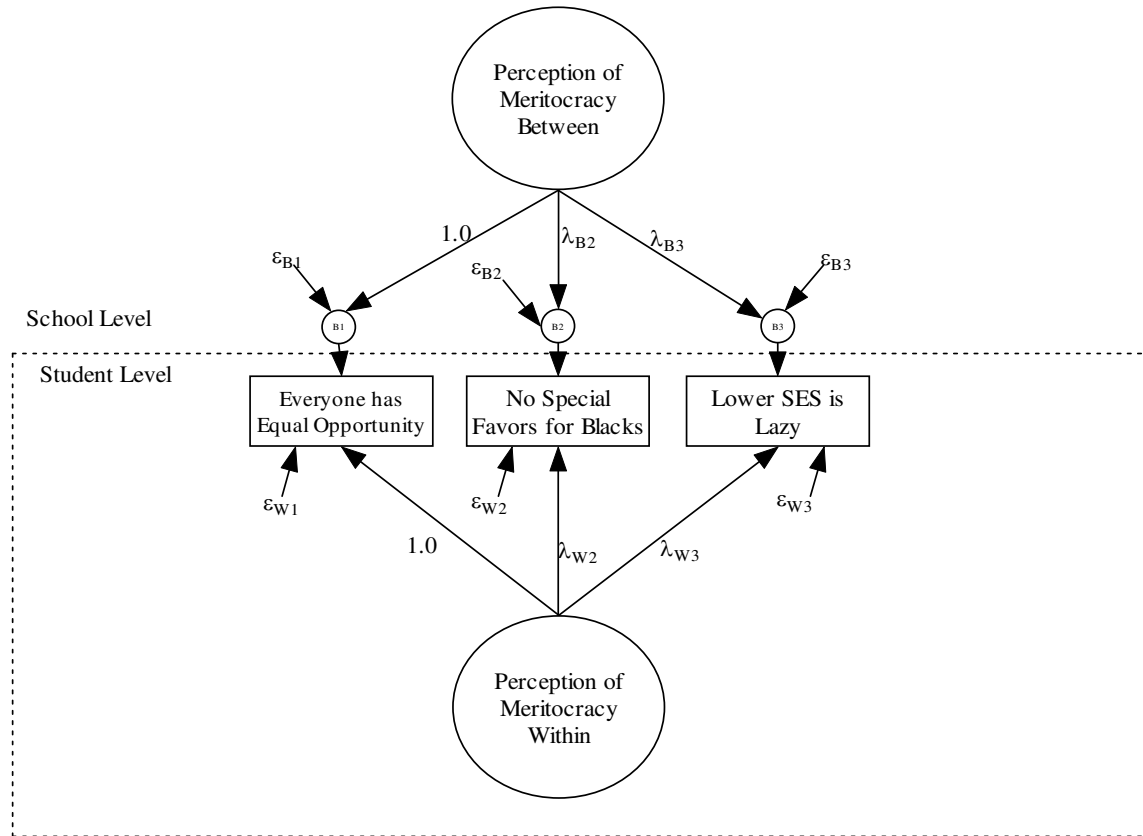
are continuous.

Figure 2. Variance Components measurement model formulation for diversity of ideas

construct. Individual-level item variances are implicitly fixed to $\pi^2/3$ by the logit function

relating the ordinally-measured items to the factor.

Figure 3. Between and Within measurement model formulation for perceptions of a

meritocracy construct. Individual-level item variances are implicitly fixed to $\pi^2/3$ by the logit

function relating the ordinally-measured items to the factor. Between-level item intercept
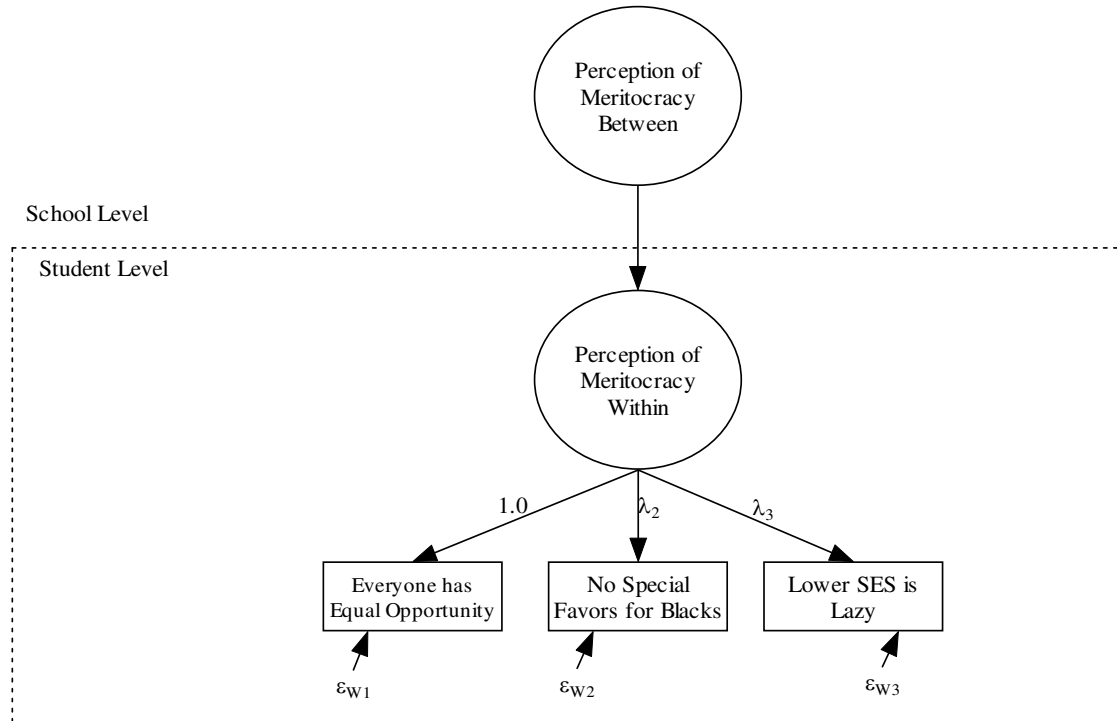
variances are continuous.

Figure 4. Variance Components measurement model formulation for perceptions of a

meritocracy construct. Individual-level item variances are implicitly fixed to $\pi^2/3$ by the logit

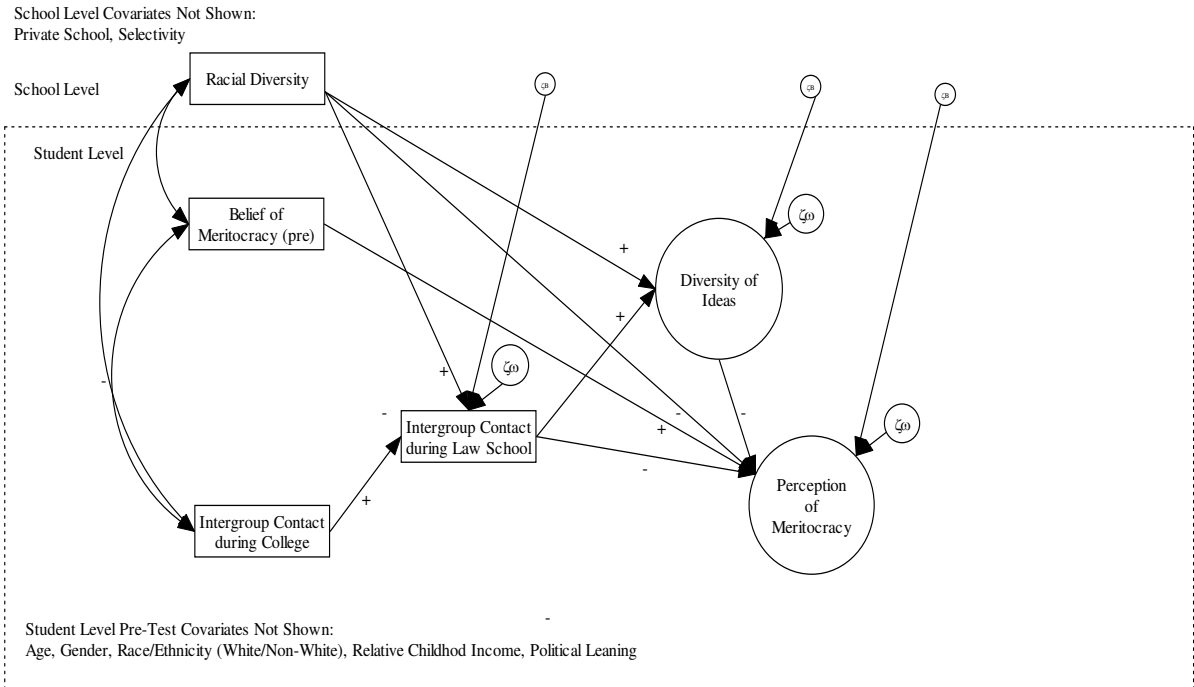function relating the ordinally-measured items to the factor.

School Level Covariates Not Shown:
Private School, Selectivity

School Level

Student Level

Racial Diversity

Belief of
Meritocracy (pre)

Diversity of
Ideas

Intergroup Contact
during Law School

Perception
of
Meritocracy

Intergroup Contact
during College

Student Level Pre-Test Covariates Not Shown:
Age, Gender, Race/Ethnicity (White/Non-White), Relative Childhod Income, Political Leaning

Figure 5. Hypothesized structural model. Plusses and minuses indicate hypothesized
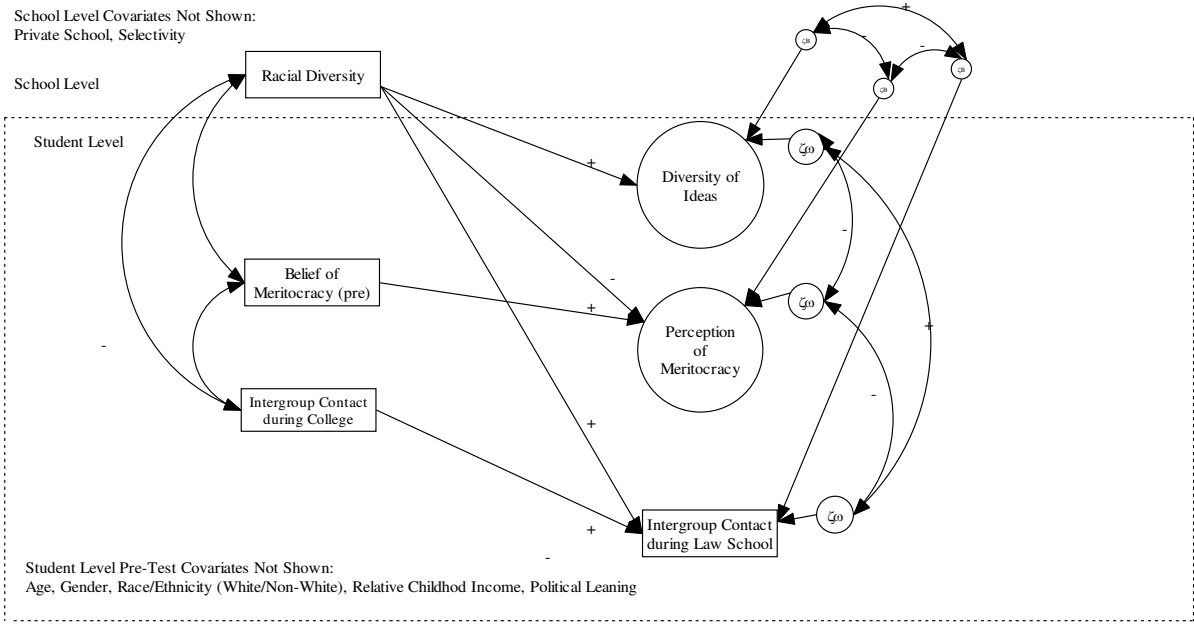
direction of effect.

Figure 6. Correlational Model- An alternative to the hypothesized model. Plusses and

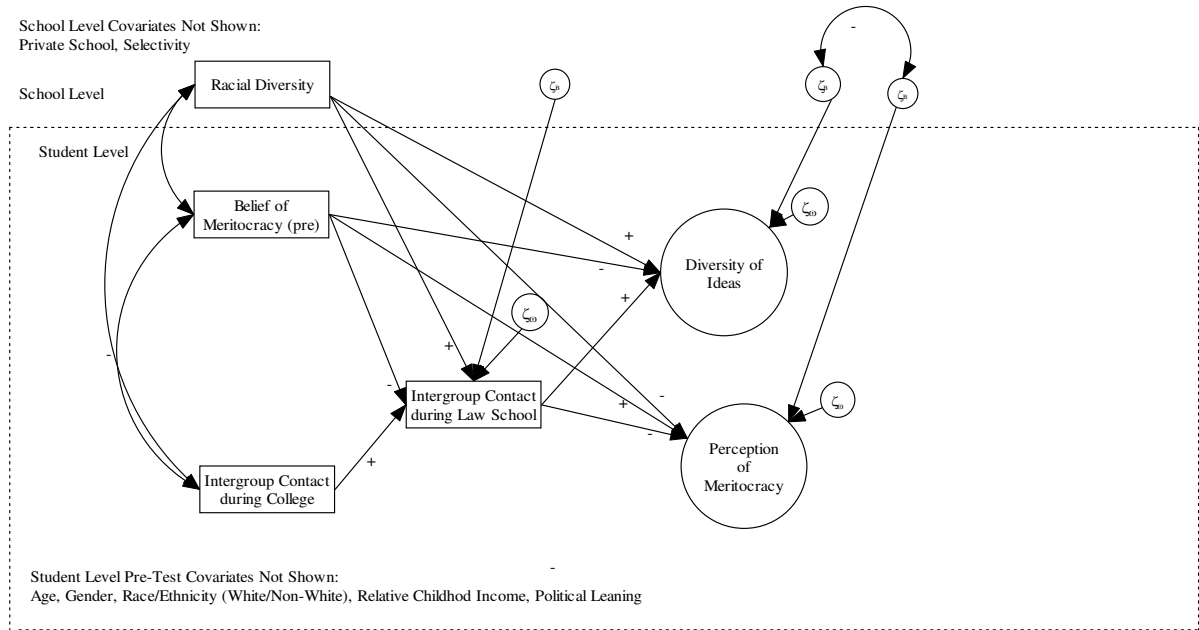minuses indicate hypothesized direction of effect.

Figure 7. Best fitting model. Statistically significant paths are drawn. Plusses and minuses indicate estimated direction of effect. Measurement models are invariant for Whites and non-Whites. Structural paths with equality constraints for Whites and non-Whites were better fitting than unequal structural paths.

References

Allen, W. R., Jayakumar, U. M., Griffin, K. A., Korn, W., & Hurtado, S. (2006, September). *Still we rise? African American higher education patterns and trends, 1971-2004.* Paper presented at the University of North Carolina, Chapel Hill Symposium on Diversity, Chapel Hill, North Carolina.

Allport, G. W. (1954). *The nature of prejudice.* New York: Addison-Wesley.

Antonio, A. L. (2001). The role of interracial interaction in the development of leadership skills and cultural knowledge and understanding. *Research in Higher Education, 42*, 593-617.

Antonio, A. L., Chang, M. L., Hakuta, K., Kenny, D. A., Levin, S. & Milem, J. F. (2004). Effects of racial diversity on complex thinking in college students. *Psychological Science, 15*, 507-10.

Arbuckle, J. L. (1996). Full information estimation in the presence of incomplete data. In G.A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling* (pp. 243–277). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Asparouhov, T. & Muthén, B. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. Proceedings of the 2007 JSM meeting in Salt Lake City, Utah, Section on Statistics in Epidemiology

Bock, R.D. & Liberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika, 35(2)*, 179-97,

Bollen, K.A. (1989). *Structural Equations with Latent Variables.* New York: John Wiley & Sons.

Bollen, K. A., Bauer, D. J., Christ, S. L., & Edwards, M. C. (in press). An overview of structural equations models and recent extensions. To appear in *Recent developments in social science statistics*. New York: Wiley.

Bowen, W.G. & Bok, D. (2000). *The shape of the river: Long-term consequences of considering race in college or university admissions.* Princeton, NJ: Princeton University Press.

Brown v. Board of Education (1954). Supreme Court of the United States.

Bucks, B. (2003). *The effects of Texas' top ten percent plan on college choice.* Report for the University of Texas at Dallas. Retrieved from: http://www.utdallas.edu/research/tsp/pdfpapers/paper34.pdf.

Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Chang, M. J. (2002). The impact of an undergraduate diversity course requirement on students' racial views and attitudes. *The Journal of General Education, 25*, 125-140.

Chang, M. J., Denson, N., Sánez, V., & Misa, K. (2006). The educational benefits of sustaining cross-racial interaction among undergraduates. *The Journal of Higher Education, 77*, 430-55.

Cook, T.D. & Campbell, D.T. (1979). *Quasi-experimentation: Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin.

Dinovitzer, R., Garth, B. G., Sander, R., Sterling, J. & Wilder, G. Z. (2004). *After the JD: First results of a national study of legal careers*. NALP and American Bar Foundations.

du Toit, S.H. & du Toit, M. (2007). Multilevel structural equation modeling. In de Leeuw, J. and Meijer, E.'s *Handbook of Multilevel Analysis*. New York: Springer.

Edwards, H. T. (2004). The journey from *Brown v. Board of Education* to *Grutter v. Bollinger*: From racial assimilation to diversity. *Michigan Law Review, 102*, 944-978.

Enders, C.K. Enders, C.K. & Bandalos, D.L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430-457.

Federico, C.M. & Sidanius, J. (2002). Racism, ideology, and affirmative action revisited: The antecedents and consequences of "principled objections" to affirmative action. *Journal of Personality and Social Psychology, 82*, 488-502.

Gottfredson, N. C., Panter, A. T., Daye, C. E., Allen, W. A., Wightman, L. F. & Deo, M. (in press). Does diversity at undergraduate institutions influence learning and democracy outcomes? *Journal of Diversity in Higher Education*.

Goldstein, H. & McDonal, R.P. (1988). A general model for the analysis of multilevel data. *Psychometrica, 53*, 435-67.

Gratz. v. Bollinger (2003). Supreme Court of the United States.

Grutter v. Bollinger (2003). Supreme Court of the United States.

Gurin, P., Dey, E. L., Hurtado, S., & Gurin, G. (2002). Diversity and higher education: Theory and impact on educational outcomes. *Harvard Educational Review, 72*, 330-66.

Holland, P.W. (1988). Causal inference, path analysis, and recursive structural equations models. *Sociological Methodology (18)*, 449-84.

Holzer, H. J., & Neumark, D. (2006). Affirmative action: What do we know? *Journal of Policy Analysis and Management, 25*, 663-90.

Horton, N.J., Lipsitz, S.R., and Parzen, M. (2003). A Potential for Bias When Rounding in Multiple Imputation. *American Statistician, 57*, 229-232.

Hoyle, R.H. & Robinson, J.C. (2004). Mediated and moderated effects in Social Psychological research: Measurement, design, and analysis issues. In C. Sansone, C.C. Morf & A.T. Panter (Eds.), *Sage Handbook of Methods in Social Psychology* (pp. 213-33). Thousand Oaks: Sage Publications.

Hurtado, S. (2005). The next generation of diversity and intergroup relations research. *Journal of Social Issues, 61*, 595-610.

Isaacs, J. (2008). The economic mobility project. *The Brookings Institution*. Retrieved from: http://economicmobility.org/reports_and_research/mobility_in_america?id=0008

Kaplan, D. & Elliott, P.R. (1997). A didactic example of multilevel structural equation modeling applicable to the study of organizations. *Structural Equation Modeling, 4*, 1-24.

Keyishian v. Board of Regents of Univ. of State of N.Y. (1967). Supreme Court of the United States.

Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods & Research, 33*, 188-229.

Lee, S.Y. & Poon, W.Y. (1998). Analysis of two-level structural equation models via EM type algorithms. *Statistica Sinica, 8*, 749-66.

Liang, J. & Bentler, P.M. (2004). An EM algorithm for fitting two-level structural equation models. *Psychometrika, 69(1)*, 101-22.

Lieberson, S. (1969). Measuring population diversity. *American Sociological Review, 34*, 850-62.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken, N.J.: Wiley.

MacCallum, R.C., Browne, M.W., Cai, L. (2007). Factor analysis models as approximations. In Cudeck, R. & MacCallum, R.C's.: *Factor Analysis at 100: Historical Developments and Future Directions*. New Jersey: Lawrence Erlbaum Associates.

MacCallum, R.C., Wegener, D.T., Uchino, B.N., & Fabrigar, L.E. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin, 114*, 185-99.

McCullagh, P.M. & Nelder, J.A. (1989). *Generalized Linear Models*. Boca Raton: Chapman & Hall/CRC.

McDonald, R.P. & Goldstein, H. (1989). Balanced versus unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology, 42*, 215-32.

Mehta, P.D. & Neale, M.C. (2005). People are variables too: Multilevel structural equations modeling. *Psychological Methods, 10*, 259-84.

Meredith v. Jefferson Co. Board of Education (2007). Supreme Court of the United States.

Millsap, R.E., & Meredith, W. (2007). Factorial invariance: Historical perspectives and new problems. In Cudeck, R. & MacCallum, R.C's.: *Factor Analysis at 100: Historical Developments and Future Directions*. New Jersey: Lawrence Erlbaum Associates.

Millsap, R. E., & Tein, J. Y. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*, 479-515.

Muthén, B. O. (1998-2004). *Mplus technical appendices*. Los Angeles, CA: Muthén & Muthén.

Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557-8.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22(3)*, 376-98.

Muthén, B. O., duToit, S.H.C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Accepted for publication in *Psychometrika*.

Muthén, B. O., Khoo, S.T., & Gustafsson, J.E. (under review). Multilevel latent variable modeling in multiple populations. *Sociological Methods & Research*.

Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus user's guide* (Fifth Edition). Los Angeles, CA: Muthén & Muthén.

Muthén, B.O. & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 25*. 267-316.

Niemaan, Y. F., & Maruyama, G. (2005). Inequities in higher education: Issues and promising practices in a world ambivalent about affirmative action. *Journal of Social Issues, 61*, 407-26.

Panter, A. T., Daye, C. E., Allen, W. A., Wightman, L. F. (2005). *Educational Diversity Project Technical Report.* Chapel Hill, NC: The Educational Diversity Project.

Parents v. Seattle (2007). Supreme Court of the United States.
Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90* 751-83.

Powers, D. A., & Ellison, C. G. (1995). Interracial contact and Black racial attitudes: The Contact Hypothesis and selectivity bias. *Social Forces, 74*, 205-226.

Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika, 69(2)*, 167-90.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics, 128*, 301-23.

Raudenbush, S.W. & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks: Sage Publications.

Raudenbush, S.W., Rowan, B.K., & Kang, S.J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to U.S. high-school data. *Journal of Educational Statistics, 16(4)*, 295-330.

Regents of the University of California v. Bakke (1978). Supreme Court of the United States.

Reyna, C., Henry, P.J., Korfmacher, W. & Tucker, A. (2005). Examining the principles in principled conservatism: The role of responsibility stereotypes as cues for deservingness in racial policy decisions. *Journal of Personality and Social Psychology, 90*, 109-28.

Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: J. Wiley & Sons.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, No. 17, **34**, Part 2.

Satorra, A. & Bentler, P. (1999). A scaled difference chi square test statistic for moment structure analysis. *UCLA Statistics Series (number 260).*

Schillings, S., & Bock, R.D. (2005). High-dimensional maximum marginal likelihood item factor analysis by adaptive quadrature. *Psychometrika, 70*, 533-55.

Shaw, E. J. (2005). *Researching the educational benefits of diversity* [Research Report]. New York: The College Board.

Shultz, M. (2007). Testing the system: Equitable law school admissions. *The Denver Business Journal.*

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equations models*. Boca Raton, FL: Chapman & Hall.

Springer, L. (1996). Applying the contact hypothesis to research on intergroup relations among college students. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, *57*, 1521.

Swygert, K., McLeod, L.D., & Thissen, D. (2001). Factor analysis for items scored in more than two categories. In D.Thissen & H. Wainer (Eds), *Test Scoring*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Thurstone, L. L. (1947). *Multiple-Factor Analysis*. Chicago: University of Chicago Press.

Whitehead, K.A. & Wittig, M.A. (2004). Discursive management of resistance to a multicultural education programme. *Qualitative Research in Psychology, 1(4)*, 267-84.

Wirth, R.J. & Edwards, M.C. (2007). Item factor analysis: Current approaches and future directions. *Psychological Methods, 12*, 58-79.

Wothke, W. (2000). Longitudinal and multi-group modeling with missing data. In T. D. Little, K. U.Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multiple group data: Practical issues, applied approaches and specific examples* (pp. 219–240). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *28*, 363-96.