SEPTEMBER 2012

CALIFORNIA
IMMIGRANT
INTEGRATION
SCORECARD

TECHNICAL REPORT

MANUEL PASTOR
RHONDA ORTIZ
VANESSA CARTER
JUSTIN SCOGGINS
ANTHONY PEREZ

California Immigrant Integration Scorecard

Technical Report

The Technical Report for the *California Immigrant Integration Scorecard* documents the methodology used to generate the data for this project. It is intended for those looking to understand the underlying data analysis.

Contents

Scanning the Field

To begin research on creating an immigrant integration scorecard, we began by scanning the field to see if any similar projects had been done. What we found were three categories of similar projects: first, profiles on immigrants that were sometimes called scorecards but did not include any scoring or ranking (i.e., Cataldo, 2008; Children Now, 2004; Fix, McHugh, Terrazas, & Laglagaron, 2008); second, scorecards that ranked jurisdictions, politicians or policies against each other (i.e. The Drum Major Institute for Public Policy, 2009); and, third, the European Migrant Integration Policy Index (MIPEX), a full-blown index on immigrant integration at the country level, that initially captured only European countries but has since expanded its reach to 31 countries. The first categories were on the scale we wanted, but the third has the scientific rigor which we sought.

A few more words on the MIPEX project based on our understanding of it at the time of this research: MIPEX's methodology is well developed and resource intensive. The indicators were designed through a series of expert consultations and later scrutinized and approved by MIPEX's Scientific Advisory Committee. More than 100 policies were researched and the project staff worked in consultation with primary and secondary national correspondents to achieve peer-review. At the time of this initial research, it included profiles of 28 countries across the following 6 strands: labor market access, family reunion, long-term residence, political participation, access to nationality, and anti-discrimination. All 140 contributing indicators are based on current laws and policies. In terms of scoring, indicators that meet the highest standards are scored with a "3," a "1" if they are the furthest from the best practice, and a "2" if in the middle. Two to nine indicators are rolled up into a single dimension and 4 dimensions make up each strand of the indicator project.

The MIPEX project addressed important questions that we found raised in the literature on scoring, particularly: "what is being measured"? Kaufmann and Kraay split indicators into two major groupings – those that measure what is on-the-books, and those that measure what is on-the-ground (2008, p. 2). Policies are a *de jure*, on-the-books measurement that the authors think are most reflective of the opinions of politicians when taken in aggregate, as in the MIPEX project. Spoonley and his colleagues (2005) more directly engage the question of what is on-the-ground in their suggestion to measure social cohesion, although in the New Zealand.

Social cohesion – which parallels "Warmth of Welcome" – can be defined as all groups having a sense of "belonging, participation, inclusion, recognition and legitimacy" (Jenson, 1998 as referenced by Spoonley et al., 2005). The authors suggest that positivist and normative indicators are needed, because "economic indicators do not provide the full story" (Spoonley et

al., 2005, p. 102 referencing Wood, 2000). In general, they aspire to nuance indicator projects: measuring social outcomes in relationship to social policy is difficult because there *is* no simple relationship. But they do push for local-level indicators because that is where social cohesion happens. Similarly, the National Neighborhood Indicator Partnership (Kahn, Kingsley, & Taylor, 2010), while offering a general indicator framework, advises that all indicators need a way to adapt to the conditions of particular places.

Park and Myers (2010) offer another set of outcome indicators.   Their article entitled, "Intergenerational Mobility in the Post-1965 Immigration Era: Estimates by an Immigrant Generation Cohort Method" measures the mobility of the second generation in comparison to the first generation. They also compare to "mainstream whites" as a way of accounting for the changing context in which the second generation is coming up. Their work was instructive in informing our "Economic Trajectory" measure, and choosing a comparison group, as will be further explained, below. Vigdor (2008) offers a similar analysis in which he uses an algorithm that quantifies how close immigrant outcomes are to native-born outcomes. Vigdor does not use an "immigrant integration" framework and assesses economic, cultural, and civic using very few indicators.

In addition, our search included scorecards outside of immigrant integration, which helped us to think about layout – an element that this review elevated to much greater importance than typical research projects.   For example, the Corporation for Enterprise Development's (cfed) 2009-2010 Assets & Opportunities Scorecard (http://scorecard.cfed.org/) used a highly accessible layout. In addition, it was the only scorecard to include measures of both outcomes and policies. For their 2010 State Energy Efficiency Scorecard, the American Council for an Energy-Efficient Economy (ACEEE) researchers determined how to rank states on energy efficiency policies when there was no national standard – heavily relying on local stakeholders. This is similar both to the MIPEX methodology as well as our context with immigrant integration wherein there is no US mandate.

As a result of these previous scorecards and literature, we came to some of the following conclusions when developing the methodology for our *California Immigrant Integration Scorecard*:

1. **Ground analysis in a strong definition of immigrant integration.** Many immigrant integration analyses are profiles, even as they might be called scorecards. Grounding our scorecard in a strong, aspirational definition provides a clear framework and direction that can be quantified. We employ a three-part definition of immigrant integration that forms the basis of our scoring: improved economic mobility for, enhanced civic participation by, and receiving society openness to immigrants.

2. **Have a clear understanding of the difference between policies and outcomes in scoring.** The literature was quite clear that scoring policies is different than scoring their effect. Using our three-part definition, we knew we would get at policy outcomes. But we were less certain about how to cover policies, themselves. And an initial search for policies across California informed us that an unbiased search on keywords was not sufficient; many policies affecting immigrants do not self-reveal and counties across the state are not consistent in their reporting. As such, we turned to our community partner – PICO California – and asked for their assistance. PICO is in-tune with policies affecting immigrants at a very local level. While we could not score these – only a larger project like MIPEX has the resources to do comprehensive policy work – we were able to incorporate the policy landscape into the context for some of the regions. For future scorecards, we are interested in deepening this analysis.

3. **Measure policy indicators in consultation with on-the-ground experts.** MIPEX and ACEEE both were able to develop high-quality policy indicators because they had reliable, on-the-ground stakeholders involved in the evaluation process. We attempted to do the same by consulting with PICO California.

4. **Use a tiered process for scoring.** Projects like MIPEX use several indicators scored up to a few dimensions, which is also suggested by some of the more theoretical literature. As a result, we decided to use that approach for each of the three definition areas. However, as we worked through the economic mobility section, we realized that it was necessary to split into two unique categories, one showing economic status and the other, economic mobility.

5. **Use tables to make indicator roll-up clear.** Looking at the different scorecards, it became clear to us that the more visually accessible scorecards would be more engaging. As a result, early in the process, we started to think about how to present data, so the scorecard would be accessible to non-researchers – organizers, business people, and policy makers are our intended audience. As a result, for example, we used standard averages instead of weighted averages across the counties – we wanted our data to be straight-forward. There's a trade-off between the simplicity of scoring and the amount of detail our scorecard offers, and we wanted to err on the side of accessibility.

6. **Choose meaningful geography.** The literature highlights the tension between too much aggregation (given that immigrant populations and integration are very local) and too little aggregation that falls short of capturing all the dynamics that affect integration – especially as regions are important geographies at which to understand the political economy of places. Choosing geography was one of the biggest initial hurdles of this project and one which we note, below.

7. **Choose measures that can be replicated:** Less from this preliminary research and more from our general knowledge of data, we wanted to be sure to choose data that would

be replicable in subsequent years should we or other researchers choose to update this scorecard. Finding publically available data at the county level that is regularly collected set strong parameters around what we could measure.

Generating Indicator Data

Having a basic understanding of where we were headed, using our three-part definition of immigrant integration, we brainstormed what types of indicators would be meaningful. During our "economic mobility" brainstorm, we determined that, both because this category had the greatest wealth of data available and because it simply made sense, it would be useful to subdivide this category into two parts: "Economic Snapshot," which focuses on how immigrants are faring in terms of current socio-economic measures, and "Economic Trajectory," which focuses on progress over time. While the term "economic mobility" sounds more consistent with the latter (Economic Trajectory), if we omitted measures of current "levels" of economic integration, we felt that the analysis would be incomplete.

The clear choice for much of the data used in our Economic Trajectory and Economic Snapshot analyses was the 5% Public Use Microdata Samples (PUMS) of the three most recent decennial censuses (1980, 1990, and 2000), and, for more recent data, the PUMS files from the American Community Survey (ACS).[1] The PUMS files are essentially samples of individual responses to the decennial censuses and the ACS, allowing for great flexibility in the derivation of measures that describe economic outcomes, and for tabulation of such measures for very specific population subgroups. While one drawback of using the PUMS files is that the geographic information included only goes down to a level known as the Public Use Microdata Area (PUMA) – an area typically containing at least 100,000 people – this level of geographic detail was deemed sufficient for our purposes.[2]

Because each year of the ACS PUMS covers about one percent of the U.S. population, while the PUMS files for the earlier years cover about five percent, three years of data (2008 through 2010) were "pooled" together to create a more comparable (and reliable) sample. While the 2008-2010 ACS sample was used to generate most of our Economic Snapshot measures (with publically-available educational testing data used for the rest), all samples, 1980 through 2008-2010 were employed in our Economic Trajectory analysis. For the latter, we chose 1980 as the

---

[1] The ACS has essentially replaced, in 2010, what was known as the "long form" in previous decennial censuses (from which the PUMS files for 2000 and earlier were generated). Both surveys (the long form and the ACS) collect detailed income, housing and employment information that is useful for our analysis. The ACS is an annual survey that began in 2000, and reached its full sample and a stabilized sampling procedure in 2005.

[2] In the 1980 PUMS, the PUMAs referred to here are called "county groups."

base year because it was the first year for which the underlying PUMA geography was sufficient for our purposes and it allows us to capture economic mobility of a large portion of immigrants who currently reside in California during their prime working years.

In choosing specific measures to include in the Economic Trajectory versus Economic Snapshot analysis, the former was somewhat limited to variables that were consistently tabulated in the various PUMS files, while variables included in the latter only had to be available in the 2008-2010 ACS. Therefore, we stuck to more traditional measures of economic status for the trajectory analysis, but expanded the set of variables in the snapshot analysis to include measures such as health coverage, social security usage, and vehicle access, among others.

For Warmth of Welcome and Civic Engagement, our brainstorm really was that – a brainstorm. There was no standard set of indicators to work from and publically-available data sets with suitable sample sizes and geographic detail do not regularly collect this type of data – particularly with regards to immigrants. For example, the November Voting and Registration Supplement to the Current Population Survey (CPS) would have been perfect, but it does not allow for detailed county-level calculations. Similarly, voter registration data is available with county-level detail (and lower), but it does not include tabulations by nativity. The same went for hate crime data which would have been very useful in understanding the warmth of welcome. As a result, we were faced with two challenges: what actually gets at warmth of welcome and civic engagement and is there data that captures that across California regions? The indicators we ended-up choosing, below, are a result of these tensions.

We also determined our geography as we were picking indicators. As noted above, the goal was to strike a balance between a level of geography that was large enough to capture regional political and economic dynamics, yet small enough so as not to lose sight of the important differences in immigrant groups and levels of integration that exist in different parts of the state. Places like San Francisco and Santa Clara have unique immigrant integration frameworks, but they also function within the context of the Bay Area region – the local economy extends beyond county lines. We wanted to go with regions that made sense as political and economic units – and that also had organizers working on immigrant integration. Overwhelmingly, this meant counties. But, we combined Riverside and San Bernardino into the Inland Empire and Alameda and Contra Costa into the East Bay – because the county couplets, to a large extent, function together and have similar trends. Given these two county groupings, we refer to the two groupings themselves the general geography used for our analysis – comprised of mostly pure counties but not entirely – simply as "regions."

The geographies we used should be considered when interpreting the Scorecard. San Francisco is much smaller than, say, Los Angeles, and immigrants are more likely to be affected by neighboring regions more so than most of the other 10 regions, with the possible exception of

Santa Clara. On the other hand, Los Angeles County is huge and immigrants are likely to live, work, and play within it, but there is a very high degree of jurisdictional complexity which complicates the political economy. In short, we acknowledge that there is no perfect geographic choice, let alone ours, and urge readers to keep in mind the tradeoffs that were part of that decision.

Finally, we chose our comparison group. Using the indicators generated from the Census Bureau, and others to be mentioned, below, we compared the immigrant/U.S.-born non-Hispanic white ratio across regions. Initially, we considered using absolute data to compare across regions – that is, simply how immigrants in different regions compared against each other – but found that the noise made by regional differences was too loud to understand what was happening in terms of immigrant integration. For example, immigrants are, of course, making less money in Fresno than in Santa Clara – but so is the native-born population. So, to level the cross-regional playing field in measuring immigrant integration, we took the ratio of each indicator for immigrant to U.S.-born non-Hispanic whites in attempt to control for regional differences. Note that for some indicators considered below, such as English-speaking ability and the naturalization rate, this methodological choice has no effect since the indicator value for natives is always one (or very close to one).[3] For others, in which there are significant regional differences in the measures for native whites (such as income-, education-, and employment-related measures) it is has an analytically valuable effect.

With our important methodological choices now laid out, we turn to a description of the actual indicators used for scoring immigrants integration in California's regions.

Economic Trajectory

In an attempt to measure economic mobility of immigrants in the various California regions studied, we followed a pseudo-cohort approach. To do so, we used the 5% Public Use Microdata Samples (PUMS) from 1980, 1990, 2000, and a pooled 2008-2010 file of the American Community Survey (ACS) microdata, all from the Integrated Public Use Microdata Series (IPUMS).[4] While the microdata samples for 1980, 1990, and 2000 all capture about five

---

[3] In some regions there are actually a small number of native-born white respondents in the survey data used who do not report speaking English at least "very well" or "only." This leads to average levels of English speaking ability among native whites in some regions that is just shy of 100% (e.g. 99.9%). However, the inclusion of this small number of less-than-perfect native-born English speakers has no discernible impacts on our results so they were left in the sample.

[4] We should note that while an "official" pooled 3-year file for the 2008-2010 ACS microdata is made available by the Census (and subsequently by IPUMS), the pooled file used for our analysis was generated by combining the three single-year IPUMS ACS files for 2008, 2009 and 2010, and then adjusting the person and household weights appropriately to yield estimates for the entire 3-year period.

percent of the total U.S. (and California) population each, a single year of the ACS microdata only covers about one percent of the overall population. Thus, we pool together three years of the ACS microdata to increase statistical reliability, reaching a sample size of about three percent of the overall population which is closer to (but not as quite as large as) the sample size in the earlier years.

For our pseudo-cohort approach, we specifically constructed six "pseudo-cohorts" for each region, defined by period of arrival in U.S. and age at time of arrival, and examined progress of each cohort over time across six socio-demographic measures, which were then combined in order to derive a final "progress score" for all immigrants in the region. While we tested a variety of approaches to measure progress, we ultimately settled on a relative approach in which progress is defined as closing of the gap between each immigrant cohort and native-born non-Hispanic whites of the same age range. A more detailed description of the analysis is provided below.

Defining Immigrant Cohorts

To define the immigrant "cohorts" that form the basis of our trajectory analysis, we initially looked at several immigrant groups defined by decade of arrival in the U.S. and age category at the time they initially appeared in the survey data. Combined, the cohorts cover all immigrants captured in the surveys used who arrived between 1970 and 2000, and were between the ages 15 and 44 when interviewed by the census (or approximately between ages 4 and 43 at the time of arrival). The cohorts are described by Table 1, below, which reports the sample size (weighted).

Table 1: Cohort Population by Sample Period

Cohort Population by Sample Period

	1980	1990	2000	2008-2010
1970's immigrant, 15-24 in 1980	477,720	504,132	498,153	409,607
1970's immigrant, 25-34 in 1980	510,300	514,629	480,862	433,254
1970's immigrant, 35-44 in 1980	216,160	218,111	185,078	174,246
1980's immigrant, 15-24 in 1990		872,255	985,018	854,159
1980's immigrant, 25-34 in 1990		963,683	981,543	849,562
1980's immigrant, 35-44 in 1990		442,976	411,682	381,719
1990's immigrant, 15-24 in 2000			741,781	714,257

			798,434	807,153
1990's immigrant, 25-34 in 2000			798,434	807,153
1990's immigrant, 35-44 in 2000			390,748	381,972

	1980	1990	2000	2008-2010
1970's immigrant, 15-24 in 1980	477,720	504,132	498,153	409,607
1970's immigrant, 25-34 in 1980	510,300	514,629	480,862	433,254
1970's immigrant, 35-44 in 1980	216,160	218,111	185,078	174,246
1980's immigrant, 15-24 in 1990		872,255	985,018	854,159
1980's immigrant, 25-34 in 1990		963,683	981,543	849,562
1980's immigrant, 35-44 in 1990		442,976	411,682	381,719
1990's immigrant, 15-24 in 2000			741,781	714,257
1990's immigrant, 25-34 in 2000			798,434	807,153
1990's immigrant, 35-44 in 2000			390,748	381,972

Note: Due to data limitations, 1970's immigrants in the 1980 sample include those who arrived in 1980 and 1980's immigrants in the 1990 sample include those who arrived in 1990.

For each cohort, we examined changes (by region) in several socio-demographic characteristics at each point in time for which they could be observed. The term "pseudo cohort" is used here because the decennial census does not track individuals (or cohorts of individuals) over time; it merely provides a cross-sectional "snapshot" of the population every ten years (or every year in the case of the American Community Survey). Thus, each pseudo cohort is constructed by selecting an age/year of arrival group in the base year survey and attempting to select the same group in each subsequent survey by increasing the age range by the number of years between surveys but keeping the period of arrival in the U.S the same. While this will certainly not select the same individuals for each cohort in each point in time, the assumption is that the individuals that are selected in each survey year for any particular cohort share certain similarities in terms of the level and trajectory of the socio-demographic characteristics examined, and thus tell us something about how that group of immigrants is progressing over time.

As Pitkin and Myers (2011, p. 266) suggest, tracking only one of these cohorts over time would do well in describing the advances made by *that* cohort, but would provide only an indication of the progress made by other immigrants during the time periods examined. Our hope is that by considering multiple cohorts we are more fully capturing the average progress made by all immigrants, and the size and share of California's immigrant population we are capturing with these cohorts suggests that this is the case. Table 2 shows the share of the state's immigrant population we are capturing in each sample year with our selected periods of arrival (in bold), while Figure 1 shows the age distribution of all immigrants in California for each of our three

selected periods of arrival, as of the first data point for each period (1980, 1990, and 2000, respectively). As can be seen, a large share of all of California's immigrants fall into the periods of arrival we consider (arriving between 1970 and 2000), and the three age groups selected cover the bulk of immigrant arriving in each decade, with the 15-24 and 25-34 being the most important in terms the size of the population captured.

Table 2: Share of Total Immigrant Population by Decade of Arrival and Sample Period

	1980	1990	2000	2008-2010
pre 1950	15%	5%	2%	1%
1950-1959	12%	6%	3%	2%
1960-1969	23%	12%	7%	6%
1970-1979	**51%**	**28%**	**18%**	**14%**
1980-1990		**50%**	**37%**	**29%**
1991-2000			**32%**	**27%**
2001 through 2008-2010 (avg)				22%
Share of all immigrants covered by decades in **bold:**	*51%*	*78%*	*88%*	*70%*

Note: Due to data limitations, 1970's immigrants in the 1980 sample include those who arrived in 1980 and 1980's immigrants in the 1990 sample include those who arrived in 1990.

Figure 1: Age Distribution of Immigrants by Period of Arrival and Sample Period
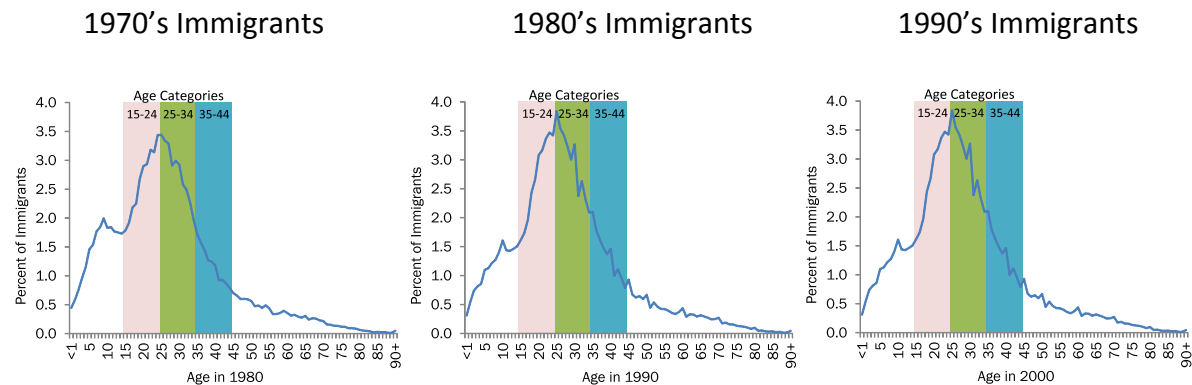
After some consideration, we decided to drop the cohorts falling in the 35-44 year old age range at their initial data point from our analysis (i.e. the cohorts represented in the third, sixth, and ninth lines in Table 1). The main reason for this is that, as seen in Figure 1, the 35-44 immigrants do not represent the "bulge" in the age distribution for each decade of arrival and thus are not likely to reflect the typical experience of immigrants from each decadal wave. Among all the immigrants arriving in California during each decade shown in Figure 1, those in

the 15-24 and 25-34 year old age ranges together comprise well over half (between 54 and 57 percent), while those in the 35-44 year old age range only account for between 12 and 14 percent. It is also that case that picking the 15-24 and 25-34 age ranges as the starting points allows our analysis to capture the most important years in terms of personal economic development – regardless of immigration experience – and not simply progress that occurs after the 35-44 age range. We thought it important to distinguish those arriving in different decades and to designate the two age cohorts for each decade of arrival (rather than one large 15-34 year old cohort), given that it is not only time in the country – but also one's age and the prevailing economic and political environment at the time of arrival – that is important to an immigrant's economic trajectory.

Measuring Economic Trajectory

To gauge the economic trajectory of immigrants in each region over time, we examined several socio-demographic indicators that are common in the literature around immigrant progress over time. These included: the rate of full-time employment (having worked at least 50 weeks and 35 hours per week during the year prior to the survey), annual personal income for full-time workers, the poverty rate (share below 150 percent of the federal poverty level), the homeownership rate (share of immigrant-cohort-headed households that own their home), English-speaking ability (the share reporting an ability of "very well" or "only"), and educational attainment (the share with at least a high school diploma or equivalent).

For each measure, we calculated progress over time in each region for each of the six designated cohorts: 1) 1970's immigrants ages 15-24 in 1980, 2) 1970's immigrants ages 25-34 in 1980, 3) 1980's immigrants ages 15-24 in 1990, 4) 1980's immigrants ages 25-34 in 1990, 5) 1990's immigrants ages 15-24 in 2000, and 6) 1990's immigrants ages 25-34 in 2000. Progress was initially figured both in an absolute and relative sense, with absolute progress defined as actual gains made in terms of the various indicators (e.g. increase in full-time employment, decline in poverty rate) and relative progress defined as the closing of gaps between immigrants and native-born non-Hispanic whites.[5]

The specific derivation of our measures of absolute and relative progress for each indicator/cohort was as follows: Absolute progress was measured by taking the value of an indicator in the base year, and considering any improvement that had occurred by the 2008-

---

[5] While we also tested the relative measures using all native-born persons as the comparison group, we felt that native-born non-Hispanic whites were a sensible benchmark as they are the group that is least likely to include large numbers of second and third generation immigrants and arguably represent "mainstream" society. In any case, the results we present around relative progress are largely the same regardless of whether we use all native-born or just native-born non-Hispanic whites as the benchmark.

2010 period. Given that all indicators (with the exception of income) are figured as rates/percentages, progress is simply measured as the percentage point increase (decrease in the case of poverty) in an indicator between the base year and last data point (2008-2010), divided by the number of years covered to make the measure of progress comparable for the different cohorts (which, of course, differ in the number of years over which progress can be examined using the various PUMS files). For income, the only indicator that is not a rate/percentage, we instead compute average annual income growth from the base year to the final data point (2008-2010).

Relative progress was figured in a similar way, but rather than using the absolute value of each indicator, the ratio of each indicator to that for native-born non-Hispanic whites falling in the same age range (the comparison group) was used. For example, for the state as a whole, the cohort of 1980's immigrants who were 25-34 years old in 1990 had a poverty rate (share below 150 percent of the poverty level) of 39 percent in 1990 while the rate for native-born non-Hispanic whites ages 25-34 was 11 percent (a ratio of 355 percent). By 2008-2010, the respective poverty rates for these two groups were 24 percent and 12 percent (a ratio of 200 percent), showing total relative progress of 155 percentage points over the entire period, or about 8 percentage points per year (155 percentage points / 19 years).

While the base year for each indicator was generally set at the first data point for which an immigrant cohort was observed in our data, there are a few exceptions. For full-time employment, the base year was set at the first data point for each cohort for which the cohort was in the 25-34 age range so that the progress captured – both in the absolute and relative sense – was that which occurred after the cohorts (and native-born non-Hispanic white comparison groups) reached prime working age. For education, a similar base year was used (again, the first data point in which a cohort was in the 25-34 age range) but with the added restriction that the base year had to be 1990 or later, as this is the first year for which high school degree (or equivalent) information was available in the PUMS (rather than simply years of school completed). Setting the base year for educational progress in this way helped to ensure that the progress captured was more comparable across the cohorts – that is, so that it is only picking up the earning of a GED among immigrants who did not have one when they entered the country for *all* cohorts rather than partly picking up that and also picking up the earning of a high school diploma at the typical age of 18 for the cohorts aged 15-24 in the base year.

Finally, given that there are six different cohorts – each with their own measure of average annual progress for each indicator – we ultimately take a weighted average across immigrant cohorts (using population in 2008-2010 as weight) to get a measure of annual average progress under each indicator for immigrants as a whole. The weights were used to allow the economic

progress of more populous immigrant cohorts to have a larger influence on the final measure of annual average progress (making the result more closely reflect the experience of the average immigrant), but the results are not very sensitive to this technical detail.

Relative or Absolute Trajectory?

With both absolute and relative measures of annual average progress figured, we were then faced with a choice: do we use the relative or absolute measures for scoring Economic Trajectory? While the choice was clear for our Economic Snapshot and other indicator categories – as they rely on *levels* from one point in time rather than *changes* over time, and we are well aware of broad regional differences that exist in the levels and the distortionary effect they can have when trying to measure immigrant integration – "going relative" was not an obvious choice for our Economic Trajectory indicator category. After all, just because levels of, say, income and homeownership are currently higher in one region than another does not necessarily mean that they have grown faster in the decades leading up to the present day.

To inform our decision, we compared the results of absolute and relative annual average progress for the different Economic Trajectory indicators across regions. Our logic: if there are not broad regional differences in trajectory in these measures over time, then both the absolute and relative measures of progress should lead to a similar ranking of regions for each (or at least some of) the indicators. Of course, our assumption here is that any broad regional differences that exist should be more or less similar (in both direction and magnitude) for immigrants and native whites, but that is an assumption we were compelled to make.

What we found suggested that controlling for regional differences in trajectory (or assuming the relative approach) was the right way to go. With regard to gains in income for full-time workers and homeownership, we saw that some regions, such as the Bay Area regions of the East Bay, San Francisco, and Santa Clara tended to do better under the absolute measures of progress, while others, such as Fresno, Los Angeles, and the Inland Empire, tended to do better under the relative measures. This suggests that the former set of regions compared more favorably to other regions when looking at overall progress of immigrants than when considering progress relative to native non-Hispanic whites of similar age, while the opposite was true of the latter set of regions; for them, the relatively slower gains of native whites made the immigrant gains look better. The pattern for these regions was somewhat different for the poverty indicator; for it, the East Bay and Santa Clara did better under the relative measure while Fresno did better under the absolute measure. For the other three indicators, each regions performance (as compared to other regions) was roughly the same under both the absolute and relative measures. Given some important differences found between the two approaches – and to be consistent with the way most of the other indicators are figured – we decided to use the relative approach.

Caveats and Limitations

Aside from the above considerations, it is important to note that the measures of economic trajectory considered are not meant to reflect as much upon how well a *region* is doing in terms of immigrant integration as how well *immigrants currently living in each region* have done in terms of integrating into society. This is because it is likely that relatively few of the immigrants in each cohort actually spent all of the time over which progress is being captured living in the same region they resided in 2008-2010. Obviously, people do move, and patterns of migration across counties and between states are not random over any period of time. An analysis of the available information in the microdata files suggests the cohorts under analysis here become more settled once they reach the 25-34 and 35-44 age range, with upwards of half of those in the 35-44 year old age group reporting residing in the same region five years prior to the survey. Some of the inter-region migration patterns are evident in the region-level data on the size of the cohort populations over time, with apparent movement out of San Francisco and Los Angeles counties and into Alameda/Contra Costa, Sacramento, San Joaquin, and Riverside/San Bernardino counties being most notable, generally, over the entire near 30-year period.

Such movement of immigrant cohorts tends to cast a shadow of uncertainty over the degree to which our measures of economic progress are telling of which counties seem to be doing better in terms of immigrant integration. For example, it is easy to imagine a scenario in which new immigrants gain a secure economic foundation in, say, Los Angeles County, and then move to the Inland Empire to buy a home – or in San Francisco and the East Bay or San Joaquin, respectively. If that were a pronounced trend, then the economic mobility that largely took place in Los Angeles could be erroneously attributed to Riverside.

It is important to recall, however, that since all of our measures of immigrant progress are figured relative to native non-Hispanic whites of similar age, the concern expressed above is not likely to be a problem if patterns of inter-regional migration are similar for the immigrant cohorts and their native non-Hispanic white comparison groups. While the latter condition is likely to be at least partly true (consider that new *and* affordable single-family homes tend to attract upwardly mobile people/renters regardless of race and nativity), it certainly cannot be taken as a given, and thus the Economic Trajectory scores for some regions are likely to be affected by inter-regional migration – particularly when the pattern of that migration differs for immigrants and native non-Hispanic whites.

To help convince ourselves that these concerns are not of grave consequence, we examined our measures of progress figured only between the first two data points (capturing gains that were made by cohorts approximately during the first two decades in the U.S.). Because the sort of economic-progress-fueled residential move described above seems less likely to occur during

the first two data points than when looking across all data points, measures of progress over this time frame would appear to be less susceptible to distortion by residential migration.

The examination had the added benefit of testing for another potential source of bias: diminishing economic progress over time. For about half of the measures we consider (homeownership, poverty and English-speaking ability) the bulk of the observed progress occurs between the first two data points. This would not be a problem if the length of time over which progress was examined were the same for each cohort, but because we can only consider progress over the first two data points (one period) for the 1990's immigrants and up to three periods of progress for 1970s immigrants, differences between regions in the distribution of immigrants across the periods of arrival could result in differences in regional progress scores. The same issues could be raised with regard to differences in the age distribution if significantly faster progress also tends to occur between, say the 15-24 and 25-34 age ranges than later in life, but the patterns found in the data did not suggest this to be a cause for concern.

In comparing the two sets of results, we found most of the results to be fairly close in terms of the relative ranking of the regions when considering any particular indicator of economic trajectory. The largest differences seemed to surface for the Fresno region with regard to full-time employment and homeownership, and for San Diego with regard to homeownership. Both of these regions scored relatively lower (compared to the other regions) along these measures when considering only progress made over the first ten years for each cohort. While we are unable to read too much into the cause of these differences, the results suggest that the issues of inter-regional mobility and asymmetric share of immigrants falling in the different cohort could be driving the scores on these indicators higher than they should be. For example, if there is a trend of somewhat better-off immigrants moving to the Fresno region specifically to buy a home and they are generally able to secure full-time employment (and the same is not true for the native non-Hispanic white comparison group), then this pattern would tend to overstate the actual progress in terms of homeownership and full-time employment. However, given that most of the regions had a similar ranking across the various indicators when considering only progress over the first 10 years for each immigrant cohort as when considering progress across all time periods, we feel that the Economic Trajectory scores reported in the Scorecard are reasonably telling of the experience of economic progress over time for immigrants.

Economic Snapshot

To generate the data for the Economic Snapshot, we used the same pooled 2008-2010 IPUMS ACS microdata that were used for the most recent data point in our Economic Trajectory analysis described above. As in that analysis, the three-year roll-up (rather than a single year of data) was used here to increase the sample size and, thus, statistical reliability. This is of

particular importance for our analysis because of the relatively low, county-level geography at which measures are calculated. Even with the three-year roll-up, the sample size was not sufficient to report data for all immigrant sub-groups in the data profiles that can be found on the back page of the regional inserts.[6] However, it was a large enough sample for the purposes of generating Economic Snapshot scores for all immigrants in each region examined.

The measures included in the Economic Snapshot are defined below in Table 3. We choose these indicators based on what data was available at the county level in the 2008-2010 microdata – we also included data from the California Department of Education (CDE). Initially, we brainstormed all the possible indicators and then grouped them by the factors influencing immigrant integration, based on our previous work (Pastor & Ortiz, 2009). Finally, we narrowed that group down to eliminate redundant indicators, as much as was feasible.

Table 3: Indicators used to measure the Economic Snapshot

Variable Name	Universe	Definition	Notes
Homeownership	Occupied households	Percent that are owner-occupied	
Rent Burden	Renter-occupied households	Percent spending 50% or more of household income on gross rent	Gross rent is the gross monthly rental cost of the housing unit, including contract rent plus additional costs for utilities (water, electricity, gas) and fuels (oil, coal, kerosene, wood, etc.). We report severe rent burden because of the high cost of living in California.

---

[6] For the data profiles noted here, and all of our analysis, we only report descriptive statistics for population groups for which there were at least 30 (unweighted) observations in the relevant microdata files. Sample sizes were far greater than this minimum cut-off the pseudo-cohorts used in our Economic Trajectory analysis. Smaller sample sizes are primarily a concern for some of the immigrant subgroups for which data was reported on the back page of the regional inserts, specifically data on self-employment and overskilled workers by race/ethnicity.

Overcrowding	Renter-occupied households	Percent with greater than 1.5 people per room	Based on the number of people in the household and the number of whole rooms used for living purposes
HS Diploma	Population ages 25-64, not in group quarters	Percent with a high school diploma or higher level of education	
Math Score	10th graders	Percent who passed the Math section of the California High School Exit Exam (CAHSEE)	We compare pass rates of English Learners to non-Hispanic white students using the 2010-2011 data series (which includes multiple test periods); Data retrieved from the CDE (http://dq.cde.ca.gov /dataquest/)
English Score	10th graders	Percent who passed the English-Language Arts section of the California High School Exit Exam (CAHSEE)	We compare pass rates of English Learners to non-Hispanic white students using the 2010-2011 data series (which includes multiple test periods); Data retrieved from the CDE (http://dq.cde.ca.gov /dataquest/)
BA or Better	Population ages 25-64, not in group quarters	Percent with a Bachelor's degree or higher level of education	

Full-Time (FT) Work	Population ages 25-64 in the labor force (employed or unemployed), not in group quarters	Percent FT workers (FT is defined as persons who usually worked at least 35 hours per week and worked at least 50 weeks during the year prior to the survey)	
Overskilled workers	Employed workers ages 25-64 with a BA or higher level of education, not in group quarters. For immigrants, the universe is further restricted to those who arrived at age 25 or older.	Percent working in an unskilled job (requiring only short-term or moderate on-the-job training, as defined by the Bureau of Labor Statistics)	The restriction to immigrants who were at least 25 years old at the time of arrival in the U.S. is intended capture only those who earned their credentials overseas. For more, see below.
Income for FT Workers	FT workers (see above definition for FT work) ages 15 and over	Median personal income	Values inflation-adjusted to 2010 dollars using the California Consumer Price Index (CPI) for all urban consumers from the California Department of Finance
Working Poor	FT workers (see above definition for FT work) ages 25-64, not in group quarters	Percent with family income below 150% of the Federal poverty level	This indicator figured only for FT workers (rather than all workers) in order to be less correlated with the percent FT workers (which is already included as a separate indicator)
Poverty	All people not in group quarters	Percent with family income below 150% of the Federal poverty level	Based on 150% of FPL because of the high cost of living in California

Health Insurance	Population ages 25-64, not in group quarters	Percent without public or private health insurance coverage	Restricted to the working-age population only because Medicaid and Medicare affect rates for youth and seniors
Cars Per Driver	Occupied households	Mean number of vehicles per person age 16 and older, by household	In a state that is still car dependent, this indicator measures access, particularly to work. By using a comparison group, differences in public transit infrastructure by region should not affect scoring.
Social Security	All people age 65 and over. For immigrants, For immigrants, the universe is further restricted to those who arrived at age 55 or earlier.	Percent receiving social security benefits	The restriction to people who were at least 55 years old at the time of arrival in the U.S. is intended capture only those with at least 10 years of eligible work (the minimum requirement to collect benefits) prior to turning age 65. See Nuschler and Siskin (2007) for more on noncitizen social security benefits.

Math and English Scores

High school math and English-Language Arts (ELA) test performance measures can capture how well-prepared English as a second language students are to enter the workforce. This data is not available by nativity but English-language skills are a proxy for immigrant students and/or the children of immigrants (who are often citizens). Our measures were generated using the 2010-2011 California High School Exit Exam (CAHSEE) research files (accessed January 2012).

Pulling the county-level data, we computed region-level Math and English-Language Arts pass rates for 10th graders for both non-Hispanic whites and English Learners (ELs).[7] For both tests, our performance measures are the ratio of EL to Non-Hispanic white pass rates.

A word on the CAHSEE: In California, all high schools students (with the exception of some students with disabilities) are required to pass the CAHSEE to graduate. The test helps ensure that all students graduate from high school with grade level skills in math, reading, and writing. The test has two major sections: math and ELA, and students must achieve passing scores on both sections in order to pass the CAHSEE. Students first take the test in the second half of 10th grade, and can retake the test in 11th and 12th grade if they fail to pass one or both sections. Pass rates are derived by computing an average pass rate (the number of students who passed the exam divided by the number of students taking the exam) across all test dates within the school year. The test is administered multiple times throughout the school year. The publicly-available dataset is accessible here: http://dq.cde.ca.gov/dataquest/.

We focused on 10th grade scores, as they represent students' initial performance on the test – creating a uniform baseline. Since the data file does not include summary information by immigrant status we used proxies: non-Hispanic whites for native-born non-Hispanic whites, and ELs for immigrants (or their children). It should be noted that while these categories are not entirely mutually exclusive, they are very close to it. By definition, an EL is a student who is not English proficient, as measured through testing, in listening, speaking, reading, and writing. A significant share of EL students are immigrants and the children of immigrants. For more on EL classification and standards, see: http://www.cde.ca.gov/sp/el/.[8]

Overskilled Workers

Because immigrants often have difficulty using their credentials from overseas, "overskilled workers" is a data point especially important to the foreign-born. We are not the first to try to quantify this measure, and draw from the Migration Policy Institute's "Uneven Progress: The Employment Pathways of Skilled Immigrants in the United States" (2008).   As a way to try to limit the sample to workers who were educated abroad, they – and we – limited the pool to immigrant workers who migrated at age 25 or older. MPI's analysis was highly nuanced –

---

[7] We use the terms English Learners (ELs) and English Language Learners (ELLs) interchangeably below.

[8] Our measure does not capture EL students who were initially classified as English fluent or later reclassified as English proficient. Their inclusion would likely paint a rosier picture, but given the lack of additional background information on former EL students (e.g., the grade in which proficiency or fluency was achieved, the nativity of the student, etc.),we instead focused on 10th grade students currently classified as EL, for more of an apples-to-apples comparison in measuring the extent to which those with the greatest needs are faring.

grouping countries by time of arrival, sending nations, three tiers of worker skill, etc. [9] While this analysis doesn't go into that sort of depth, we do take their approach of matching ACS occupational codes to the Bureau of Labor Statistics (BLS) educational/training categories.

To match the BLS data with the ACS data, we started by downloading the BLS training level data for 2008, the initial year of our analysis (Formerly, here: http://www.bls.gov/emp/ep_table_106.htm, now here: http://www.bls.gov/emp/ep_education_training_system.htm).   The BLS uses Standard Occupational Classification (SOC) System codes for training level data. To read the BLS data into the ACS data, we had to create a match file. Based on the 2000-2004 ACS OCCSOC code (it is the most complete version, as compared to the 2005-2007 which has collapsed categories to ensure confidentiality, and at that time those were the most current options), we matched the occupation title and training code to the ACS data. While the vast majority matched correctly, there were a few cases in which no training levels were returned. In several instances, this was because the BLS only assigned values to subcategories, not umbrella categories. But to be thorough, we averaged the subcategories and assigned that value to the umbrella category (e.g., Marketing and Sales managers, code 112020). In other instances, titles or codes were slightly off. For example, BLS code 131199 and the ACS OCCSOC code 1311XX (a generic category) appeared to be a match but the titles were slightly different – "Other Business Operation Specialists" and "Business operations specialists, all other." We called it a match and assigned the corresponding value. This hand-matching happened in a handful of cases. We excluded military occupations from the analysis.   Finally, matching OCCSOC codes 194051 and 1940XX to BLS codes 194051, 194090 and its subcategories, and 194061 was tricky and did not follow any of the above patterns and we had to make some informed guesses about training values. For the full detail on this matching, please contact CSII.

We also wanted to have a full set of BLS training level data for the 2005-2007 ACS OCCSOC codes. The file we used to match the BLS data into the 2000-2004 OCCSOC codes was actually a crosswalk available through IPUMS (http://usa.ipums.org/usa/volii/acs_occtooccsoc.shtml; Retrieved 16 Sept 2011) that included the following variables: ACS OCC code (2-digit), 2005-2007 ACS OCC code (2-digit), ACS OCCSOC code (6-digit), 2005-2007 ACS OCCSOC code (6-digit), and the occupation name. So, having already assigned training level data to the 2000-2004

---

[9] "Unskilled occupations" require no more than modest on-the-job training (e.g., construction laborers, customer-service representatives, child-care workers, house cleaners and maids, file clerks). Skilled technical occupations typically employ workers with long-term on-the-job training, vocational training, or associate's degrees (e.g. carpenters, electricians, chefs and head cooks, massage therapists, real estate brokers). High-skilled occupations require at least a bachelor's degree (e.g. scientists and engineers, doctors, financial managers, postsecondary teachers)." (Batalova & Fix, 2008, p. 13)

data and having the 2000-2004 already crosswalked to the 2005-2007 data, the majority of data was already matched. However, there were a few inconsistencies. In some cases IPUMS lumped several 2005-2007 codes into one, like with 1110XX. We simply assigned '1110XX' to both corresponding codes and deleted the lumped version. In this way, we had more detailed training data. We only deleted the lumped code when it was completely distributed between more detailed categories. Some lumped categories included left-over codes that could not be broken back out. We left them in the lumped category (i.e. 1520XX, "Miscellaneous…"). For code 19040XX, we broke it down as much as possible. We were able to move everything out of the lumped category, except social science research assistants. This class of worker simply was not broken out in the 2005-2007 ACS OCCSOC code. We retitled the lumped category and assigned the value for that occupation from the BLS training data. 1940XX is now three line items in the 2005-2007 ACS OCCSOCC column; the lumped line item now reflects only "social science research assistants." For 453000 ("Hunters and Trappers"), there was no associated training-level value in the BLS. We assigned it the "Fisher and related fishing workers" value. As with before, we excluded all military occupations.

Nearly there and now with a complete and reliable crosswalk between 2000-2004 and 2005-2007 OCCSOC codes and BLS training level data, we used a pivot table to return single training values for each OCCSOC codes, based on simple averages, for both periods of time. Remember, that sometimes we had to duplicate OCCSOC codes in order to get reliable matches with the BLS data.   Both pivot tables were then converted into SPSS .sav files with OCCSOC values and rounded and unrounded training values (because averages were used in creating the crosswalk).   In SPSS, we then matched in this occupational training data and were able to do an analysis of "overskilled" workers. As per the MPI definitions, we considered overskilled workers to be those with Bachelor degrees or higher working in jobs that required no more than modest on-the-job training.

Warmth of Welcome

This category takes seriously the understanding that immigrants contribute to the strength of their region – and so measures if the region views them favorably and worth the investment. Data on this measure was hard to come by. For example, we had really hoped to include hate crime data, but it is not collected by crimes against immigrants. Further, warmth of welcome data is not the type of data typically collected by government agencies.

Media Score

To capture the popular regional narrative on immigrants, we performed a media analysis. We conducted a search and qualitative analysis of the content of a number of print and digital

media outlets that serve both general and specific audiences across the 10 regions in the Scorecard. Using MondoTimes and ABYZ News Links, we identified the news outlets serving each region, the type of audiences they serve and the number of residents they reach (circulation numbers). Based on this information, we chose to analyze the news media outlets with the highest circulation and serving the following three audiences: general population, ethnic groups, and university students.[10] Thus, a minimum of three different media outlets were examined for each region.

In order to collect the articles from each source, when possible, publications were searched using Lexis Nexis. Using the search term "immigration" ("*inmigracion*" for Spanish news sources) and narrowing our search to all articles produced within the last 12 months of the start of the project, we were able to generate a set of articles that had been produced by each news outlet and record the total. In other instances, we searched the archives available on the outlet's websites, using the same search terms and keeping the same time span (ie: searching the Los Angeles Times' archive for immigration articles published in the last 12 months).

Based on the sets that were generated from each search, we tried to gather 2 articles published per month—for newspapers serving the general audience—for a sample size of 24 articles. Since newspapers serving ethnic communities or universities produce fewer stories each month, we had to use our judgment and try to collect as many stories published within the last year to assess the way these outlets reported on immigrant issues. Each article was then read and examined for: tone, topic and references/spokespeople. Tone was assigned as negative, neutral or positive based on the attitude that the author took in the article through the use of syntax, imagery or vivid appeals and information omitted or used. Topic refers to the specific issue—under the immigration umbrella—the article covers. And spokespeople refers to the people that the author used as references to talk further about the issue, or to present different sides to the issue.

In order to comprehensively score each region's narrative on immigrants, two factors were weighted and brought together into a final score.   The first factor captures relevance and reach. For each source, the number of immigration articles published within the last 12 months was multiplied by their average circulation. The products (circulation * articles published) of all 3 sources were added to get a regional total. Those products were then divided by the regional total to get a percentage for each source that represents their relevance in comparison to all other news sources in that region – the thinking is that the news sources that produce more articles and have a larger audience are more likely to influence and represent the opinions of

---

[10] The San Francisco region has two large newspapers, with significantly similar circulation numbers, which serve the general audience. For the purpose of this study, both of them were included in our analysis.

local residents and policymakers on immigration and immigrants. An example of this weighting is provided in Table 4 below.

Table 4: Relevance and Reach Weighting in the Sacramento region

Circulation Information (for weighting)			
Source	Sacramento Bee	El Hispano	The State Hornet
Number of articles	268	0	5
Average circulation	279,032	20,000	8,000
# of articles * circulation	74780576	0	40,000
Regional weight	0.9995	0	0.0005

The second factor captures tone. As mentioned above, each article was categorized by tone. The share of positive, neutral and negative articles for each source was calculated, each multiplied by its weight (3 for positive, 2 for neutral, 1 for negative), and then added to get an individual score for each media source. For example, if a source only published positive articles, then it would have 100% positive articles multiplied by the tone score of 3 to produce an overall score of 3. The relevance and reach score was then multiplied by the tone score for each source and the scores for each news source were added to get an overall regional score. The individual regional scores were used to rank the regions against each other on a 1-5 scale. The details of that process are noted below.

Academic Performance Index

The ability of schools to educate students with limited English indicates their capacity to invest in newcomers. In order to evaluate this, we analyzed elementary school Academic Performance Index (API) scores of EL students relative to non-Hispanic white students. The API measures the annual academic performance of schools in California based on a number of statewide standardized tests. The API score is a single number ranging from 200 to 1000, with 800 being the target statewide and highly-performing schools aiming to maintain a score at or above 800. Due to frequent changes in the scoring methodology from one year to the next, two scores are released each year: a "base" score that is calculated using the most current methodology and inputs, and a "growth" score that is calculated using the methodology and inputs from the previous year, so that changes in performance can be consistently evaluated. For more, see: http://www.cde.ca.gov/ta/ac/ap/documents/infoguide12.pdf.

Our measures were generated using the 2010 API base score data for elementary schools (retrieved from http://www.cde.ca.gov/ta/ac/ap/, accessed September 2011). We chose the base score as it reflects the most current methodology, and looked at levels (rather than

changes from the previous year) to be consistent with the way other indicators in the Warmth of Welcome category are figured. To compute regional API scores for ELs and non-Hispanic whites, we took a weighted average of their respective scores across all elementary schools in the region (using the number EL and non-Hispanic white students, respectively, as weights). For more on EL classification, see the "Math and English Scores" section above.

The universe reflected in our API measures differs from similar summary-level measures available online (see http://dq.cde.ca.gov/dataquest/) in a couple of ways. First, the grades which are considered elementary differ – in our measures, which use CDE's school-level dataset, the "elementary school" classification (found in the "school type" code) generally captures students in kindergarten through fifth grade.[11] In contrast, the API measures found in summary reports and tables consider grades two through six elementary, and are generated using grade-level information from a person-level dataset – data which is not available for public download.

The second difference is related to the overall universe of students reflected in each dataset. The population captured in the statewide summary tables includes the test scores of every student in the state regardless of the student's California Basic Educational Data System (CBEDS) status (for more, see: http://www.cde.ca.gov/ds/dc/cb/). However, a school's API (and our measure) is calculated based on the valid test scores of only those students who enrolled during the CBEDS period and continuously enrolled until the first day of the Standardized Testing and Reporting (STAR) tests are given (for more, see: http://www.cde.ca.gov/ta/tg/sr/).[12]

Immigrant-Serving Organizations

Our "immigrant-serving organizations" indicator captures each region's pool of immigrant-serving nonprofits relative to the size of its non-citizen immigrant population. Immigrant-serving organizations play a pivotal role – especially for undocumented and non-citizen immigrants – in ensuring that immigrants are receiving the services necessary to be healthy, productive, engaged and informed members of society.

To create this measure we used Guidestar (http://www2.guidestar.org/), which includes a searchable, online, nonprofit database.   While the Guidestar database contains a wealth of information, not all of it is available for free. Given our desire to create a transparent and easily replicable scorecard methodology (as well as our own budgetary concerns!) we drew from the

---

[11] However, it should be noted that not all elementary schools conform to this grade range (e.g., some elementary and middle schools may include students in grades nine through 12).

[12] These caveats were outlined by Renyi Liu of the Measurement and Accountability Reporting Division of the CDE in an email correspondence, January 2012.

no-fee data. One of the limitations of the free-version is the lack of detailed firm information. For example, because we have no firm size data, each nonprofit is treated equally in our analysis. Information on firm income is available, but given its definition, it cannot easily be used as a proxy for size.

To identify immigrant-serving organizations, we queried on category "P84" (Ethnic and Immigrant Centers, Services). This category pertains to organizations that provide or coordinate a wide variety of programs and services that are structured to meet the social, educational, economic, recreational and other needs of specific ethnic and/or immigrant groups in ways that are culturally appropriate.[13] As useful as this categorization is, it does not include political or advocacy organizations and may also miss organizations that have broader missions but serve many immigrants along the way. Nonetheless, it does give a good relative picture when comparing regions against each other.

With 457 records in hand we went about creating a single score for each county. To match organizations to counties, we used geographic information systems (GIS) software to map the data by zip code and then spatially join the zip codes to the county. In order to normalize the data, we divided the number of organizations per county by the number of non-citizen immigrants (in 1000s) – a data point generated using the pooled 2008-2010 IPUMS ACS dataset described above.

Civic Infrastructure for Naturalization

This indicator is meant to capture the quality and availability of civic infrastructure to facilitate the naturalization process. By civic infrastructure, we intend to include the different types of "bricks and mortar" infrastructure, such as government offices that process naturalization applications and community-based organizations or adult schools that provide education and promote and facilitate naturalization, as well as what might be referred to as "cultural infrastructure," such as employers that encourage their workers to naturalize and immigrant social networks that more effectively encourage naturalization and produce greater naturalization rates. In trying to come up with a quantitative indicator that captures these different sorts of civic infrastructure for naturalization, however, we were faced with a serious lack of data. Therefore, we decided to generate an estimate of such infrastructure using an admittedly indirect – yet statistically sophisticated – approach.

---

[13] Definition from the Urban Institute's National Center for Charitable Statistics. For more, see: http://nccsdataweb.urban.org/PubApps/nteeSearch.php?returnElement=&popup=0&gQry=P84&codeType=NTEE

To estimate the quality of civic infrastructure for naturalization, we developed a regression model to predict county-level naturalization rates across all counties in US, controlling for many important factors that are strongly correlated with naturalization rates, such as the recency-of-arrival- and country-of-origin composition of each county and state dummy variables to control for any state-specific effects. We then calculated the difference between a county's *actual* naturalization rate and its *predicted* naturalization rate, and used this difference as an indicator of civic infrastructure for naturalization. The indicator essentially captures "over performance" in naturalization – that is, the extent to which the naturalization rate in a county is higher than would be expected given the composition of its immigrant population.

The logic behind this approach is as follows: the difference between the actual and predicted naturalization rate for a county (whether positive or negative) is essentially the portion of the naturalization rate that cannot be explained by characteristics of the immigrants in the county (or other factors included in the regression model), and is assumed to include the effect of other factors that may be important, such as civic infrastructure, as well as a random error term. While this approach is relatively crude and provides merely a rough estimate of what we are trying to measure, we take some comfort in the fact that the characteristics included on the right-hand-side of the regression model are highly predictive of county-level naturalization rates, explaining 87 percent of the total variation in county-level naturalization rates, based on the adjusted r-squared of the regression. This means that the unexplained portion is relatively small, and since civic infrastructure is assumed to be an important part of the unexplained variation, it is likely an important component of the indicator we use to approximate it.

The data on which the regression model was based was obtained via a Freedom of Information Act (FOIA) request to the Office of Immigration Statistics (OIS), which is a part of the U.S. Department of Homeland Security (DHS). It includes counts of all Legal Permanent Residents (LPRs) who attained LPR status between 1985 and 2010, by county, period of status attainment (1985-1991, 1992-1998, 1999-2005, and 2005-2010), country of origin (for roughly the top 20 countries of origin for immigrants in the U.S. in 2010), and whether or not citizenship had been obtained as of 2010. Before analyzing the data, basic adjustments were made for mortality, derivative citizenship and emigration (similar to those described in Rytina, 2011).

The "naturalization rate," which is used both as the dependent variable in the regression model and as an indicator in the Civic Engagement category described below, was figured only for LPRs who attained status between 1985 and 2005 (since those attaining status later would likely not be eligible to naturalize by 2010 given the typical five-year residency requirement). It was specifically calculated as the share of all such LPRs that had naturalized as of 2010.

The regression model itself was restricted to counties with at least 2,000 LPRs who attained status between 1985 and 2005, and modeled the naturalization rate as a function of the shares

of longer-term LPRs (those attaining status in 1985-1991 and 1992-1998, respectively) as they are more likely to have naturalized as of 2010, the shares of LPRs from specific countries and broad regions of origin (since naturalization rates vary dramatically by country/region of origin), and dummy variables for the top six LPR-receiving states based on the LPR sample considered (California, New York, Florida, Texas, New Jersey, and Illinois) as they are suspected to have more developed immigrant-serving infrastructure which may affect naturalization rates. To be consistent with the geography of the regions included in our analysis, data for Riverside and San Bernardino counties were combined into a single observation, and the same was done for Alameda and Contra Costa counties.

ELL Supply Relative to Need

The more a county sees their immigrants as contributors, the more they will see the demand for English language learning and provide classes. To measure the supply of English language instruction relative to the demand, we drew on an assessment of English language instruction in California (Fix, Laglagaron, Terrazas, & McHugh, 2007)[14],in which the Migration Policy Institute (MPI) estimated both the number of individuals who require adult ESL instruction and the provision of ESL instruction by county – as well as the gap between the two. We drew on data indicating the estimated share of adult ESL need that was supplied directly from this report for single county regions, and calculated a weighted average for multi-county regions. While immigrants make up a large share of the ESL instruction need, a small share is comprised of native-born who also lack the English skills necessary to participate fully socially and economically.

In order to estimate the supply of ESL courses by county, MPI analyzed 2000-2006 data from California's two primary providers of adult English instruction: adult schools and community colleges.[15] [16] Although the two systems may informally coordinate their services on the local

---

[14] The ESL "supply versus need" framework is developed in Fix et al. (2007), but we draw our data from an extract of a related 2008 report, "An Assessment of the English Language Instruction Need and Supply in California," commissioned by the Grantmakers Concerned with Immigrants and Refugees' California Immigrant Integration Initiative and researched by the Migration Policy Institute's National Center on Immigrant Integration Policy. That data was available in a draft document formerly found on the web (now stored in our files), and a report has been published, although without data on Southern California counties, here: http://www.gcir.org/system/files/GCIR_ESLissuebrief_web.pdf.

[15] The ESL supply methodology described here is drawn directly from a February 2012 email conversation with Michael Fix of the Migration Policy Institute.

[16] The estimates of ESL supply are based MPI tabulations of data instruction hours data from the California Community College Chancellor's Office and the California Department of Education, Office of Adult Education, 2000-2006.

level, there is no central office or set of processes to coordinate the two systems statewide. MPI estimated the provision of services for each system separately. The approach for each included:

1. Estimating the Supply of ESL Instruction Provided By Adult Schools

    Adult schools are operated by local school districts with administrative and technical guidance from the CDE. Each adult school course is required to report its daily attendance to the state office. The average daily attendance (ADA) record is then used by the state to disburse funds to the districts.[17] Since ADA can be disaggregated by program area it was used as the basis for calculating the hours of adult ESL instruction provided by the adult school system.[18] In order to focus specifically on ESL courses, MPI multiplied the total adult education ADA by the share of adult school enrollment in ESL courses as reported in the CDE's annual fact book. The resulting figure provides a reasonably reliable estimate of the number of hours of ESL instruction provided by the adult education system.[19]

2. Estimating the Supply of ESL Instruction Provided By Community Colleges

    Community colleges are publicly-supported and locally-oriented educational institutions that are loosely organized statewide by the Sacramento-based California Community College Chancellor's Office (CCCCO), to which attendance and enrollment data are reported. Community colleges report the number of full-time equivalent students (FTES).[20] Community colleges offer both credit and non-credit ESL courses and report the number of FTES for both. The concept of a FTES allows the colleges to combine "shares" of part-time students with full-time students for reporting purposes. The FTES calculation is based either

---

[17] ADA is calculated by dividing total attendance by 525, which is the number of instructional hours provided per student per course on a yearly basis. There are 180 days in a school year and adult education courses typically involve three hours of instruction per day, minus approximately 25 hours to account for holidays and in-service days.

[18] While ADA can be disaggregated by program area (i.e., adult education, special education, etc.) it cannot then be disaggregated by course (i.e., ESL, Adult Basic Education, etc.).

[19] For example, according to the CDE Factbook, 41.6 percent of total adult school enrollment statewide in 2005-06 was in ESL courses. The share of adult education enrollment in ESL courses between 2000 and 2006 varied from 40.3 to 43.3 percent of total enrollment. Experts at the CDE suggest that this may underestimate the actual share of adult enrollment in ESL due to some reclassification of adult ESL students into other course areas - notably adult basic education (ABE) and citizenship courses; however, we use the official figure in our analysis.

[20] Like ADA units, FTES are equal 525 hours of classroom instruction.

on daily attendance records or, in certain courses, enrollment at a predetermined "census point" - usually one-fifth through the course.[21]

Complications arise, however in trying to develop county-level estimates of ESL instruction provided by community colleges. Due to the inconsistent and often overlapping boundaries between school districts, community college districts, and counties, MPI assigned each community college to a specific county. They did this by identifying the city and county in which the college's main campus was located (excluding satellite or alternative campuses). However - particularly in the densely populated metropolitan areas surrounding Los Angeles and San Francisco and in rural regions where the community college infrastructure may by sparse - the county in which a community college is located may not necessarily correlate with the counties from which it draws its student population. Similarly, where counties are small and compact, the data may suggest that the community college system provides no ESL instruction, when in fact, it is probable that the county's population is being served by colleges in bordering counties.

Turning to the demand side, the estimates of ESL need of adult Legal Permanent Residents (LPRs), unauthorized immigrants, refugees, natives, and naturalized citizens in California were calculated by analyzing 2000 Census data.[22] The key variables used in MPI's estimates of need included:

- *Population size by group* (e.g. of LPRs, unauthorized immigrants, etc.) – among immigrants, provides an overall snapshot of potential universe of English learners

- *Age* – indicates the type of English language instruction (e.g. adult instruction) which is needed

- *Educational attainment* – in part, measures an immigrant's ability to acquire or have acquired English skills

- *Time spent in the U.S. (for immigrants)* – indicates an immigrant's potential English-learning opportunities

Taken together, MPI analyzed the aggregate adult ESL needs as well as the demand by education levels, age, and time spent in the U.S, in order to further understand the specific types of services which were needed.

---

[21] Tom Nobert, Information System Specialist, Division of Management Information Services, California Community College Chancellor's Office, interview by Migration Policy Institute, interview by telephone on November 13, 2007.

[22] MPI tabulations of 2000 Census data with imputations of legal status by the Urban Institute.

Civic Engagement

This category captures the extent to which immigrants are able to engage in government and social processes that affect both their personal and community-wide well-being. Of the four categories, we were able to find the least data, here. While there is good data on voting, it is not collected based on immigrant status.

Linguistic Integration

Linguistic integration is the opposite of linguistic isolation – the latter is reported on the back page table for each region. A linguistically integrated household is one where there is at least one person 14 years or older who speaks English only or "very well." We included this measure because language can affect the ability of immigrants to fully participate in civic life.

Naturalization Rate

To estimate the naturalization rate for immigrants in each region, we utilized the same data that was used to estimate civic infrastructure for naturalization, described above. See the description of that indicator for how naturalization rates are calculated.

Scoring Methodology

As was explicitly acknowledged in the Economic Trajectory section and implicitly referenced elsewhere, most indicators required a comparison group in order to be scored. Once it was decided that a relative measure (that is, scoring based on gaps between immigrants and the comparison group) was better than an absolute measure (scoring based on immigrant data only), the relative approach was applied to all indicator categories where possible and appropriate. In most cases, the same comparison group was used for all indicators (U.S.-born non-Hispanic whites). Given data constraints or a focus on the children of immigrants, there were some exceptions: our Math and English scores and API scores compared English learner students to non-Hispanic white students (rather than U.S.-born non-Hispanic whites). No comparison groups were used for Media Score, Coverage of Immigrants Serving Organizations, the Civic Infrastructure for Naturalization, ELL Supply Relative to Need, or the Naturalization Rate – so these could be considered absolute measures. Using a relative score is also useful in that is avoids any claim of any region having "made it" – that is, scoring perfectly on immigrant integration.

For each indicator, figured as either a ratio of the value for immigrants in the region to the comparison group, or as an absolute value in the few cases noted above, we applied a 1 to 5 scoring system. Using a common normalization technique to make "apples-to-apples" comparisons between variables with different scales and distributions, the mean and standard deviation of each indicator across all ten regions was calculated and referenced for scoring each

region under each indicator. Specifically, the indicator value for each region was scored based on how many standard deviations above or below the mean across all regions it fell (see Table 5 for details). Under this scheme, "5" indicates higher levels of integration while "1" indicates lower levels.

Table 5: Scoring Assignment*

Score	Description
5	*Greater than one Standard Deviation (St. Dev.) above mean*
4	*Greater than 0.5 but less than or equal to 1 St. Dev. above mean*
3	*Within 0.5 St. Dev. of mean(including 0.5 above, but not 0.5 below)*
2	*Greater than or equal to 0.5 but less than 1 St. Dev. below mean*
1	*More than or equal to one St. Dev. below mean*

* For some indicators (e.g. Poverty), the scale worked in reverse, where, for example, a "5" would be 1 St. Dev. below the mean instead of above.

Once each of the 28 indicators were scored in this way, category scores for Economic Snapshot, Economic Trajectory, Warmth of Welcome, and Civic Engagement were computed as simple averages across all the scores for indicators included in each category. The category scores were then used to compute an overall score for each region as the simple average across the four category scores. This approach has implications around the weight of the indicators. For example, while the Economic Snapshot considers 15 indicators in generating a category score, Civic Engagement only considers two – and so the Civic Engagement indicators carry more weight in determining the category scores (and hence the overall scores). Finally, the importance placed on economic integration in generating the final score is clear: there are two economic categories, and since each of the four categories is given equal weight in the final score, economics account for 50% of it, while and Warmth of Welcome and Civic Engagement account for 25% each. This importance placed on economics was intentional, reflecting both our judgment that reasonably solid economic footing is often supportive of increased civic engagement (making it more important), and that the data sources and indicators underlying the economic categories are more reliable and established in the field (and thus deserve more weight).

With scores assigned by indicator, category and overall, we also ranked the regions. For ease of analysis (and to not overstate the accuracy of our scores!), we rounded all scores to one decimal point which is how they are reported in the document. We should note that while there were some regions that tied with others in their ranking before the rounding, a few more tied after the rounding – and we report them as ties nonetheless. Table 6 indicates rankings of each region in overall and category scores, with ties shaded in grey.

Table 6: Regional Scores and Ranks

	Overall Score		Economic Snapshot		Economic Trajectory		Warmth of Welcome		Civic Engagement	
	Score	Rank	Score	Rank	Score	Rank	Score	Rank	Score	Rank
Santa Clara	4.0	1	3.8	1	3.7	3	3.4	3	5.0	1
East Bay	3.4	2	3.3	5	3.0	4	3.2	5	4.0	2
San Diego	3.2	3	3.5	2	3.0	4	2.8	6	3.5	3
Orange	3.1	4	2.4	8	3.8	2	2.8	6	3.5	3
Sacramento	3.1	4	3.5	2	2.7	8	3.4	3	3.0	5
San Francisco	3.1	4	3.4	4	1.7	10	4.4	1	3.0	5
Inland Empire	2.7	7	3.2	6	2.2	9	2.4	8	3.0	5
Los Angeles	2.6	8	1.9	10	3.0	4	3.6	2	2.0	8
San Joaquin	2.6	8	2.7	7	4.0	1	2.2	9	1.5	9
Fresno	2.0	10	2.1	9	2.8	7	2.0	10	1.0	10

Back Page Table and Graphs

In addition to scoring immigrant integration, contextual data is found on the back page. Most of this data is from the same pooled 2008-2010 file of the ACS PUMS data that was used for the majority of the analysis in the Scorecard. There are some exceptions: for the unauthorized status, we used the pooled 2008-2010 ACS PUMS file in conjunction with an estimating technique (and regression coefficient estimates) provided by Enrico Marcelli at San Diego State University, and for data on LPRs we used the data obtained from the Office of Immigration Statistics via a FOIA request, as described above in the "Civic Infrastructure for Naturalization" section.

Most of this data is straight-forward, and there are footnotes where it is not. The following are a few points needing elaboration:

- **Linguistic Isolation**: The reverse of linguistic integration, this shows the proportion of households in which no person age 14 or older speaks English only or "very well."

- **Unauthorized Status**: To generate this data, we used an equation developed by Enrico Marcelli at San Diego State University that estimates who is undocumented. This estimation procedure is only available for adult Latinos. For a recent description of the approach, see Marcelli and Lowell (2005).

- **Sanctuary City**: Initially we had hoped to include this data in the Warmth of Welcome category, but it did not lend itself to scoring as an all or nothing data point. Moreover, sanctuary city status is applied with no clear uniformity – sometimes cities declare themselves as such, sometimes not – and there is no associated legal meaning.

  For our list, we pulled data from both a Congressional Research Services (CRS) document (Seghetti, Vina, & Ester, 2006) which is widely referred to by many pro- and anti- immigrant groups who write on this topic, and which we supplemented by listings from an anti-immigrant website (http://ojjpac.org/sanctuary.asp). Our logic: if anti-immigrant groups were noticing these cities, they must be doing something right by immigrants! If a city was noted on the website, but not in the CRS report, additional internet searching was done.

- **Self-Employment**: We included this variable on the back page instead scoring it because its meaning remains unclear. Portes and Zhou (1996) highlight that entrepreneurship can mean both exclusions from traditional businesses or that there are opportunities for immigrants (who are able to start businesses in that region. Portes and Zhou (1996) ultimately come down on the side of higher rates of entrepreneurship being a positive indicator of integration, but others point to limited language ability (Mora & Dávila, 2005; Portes & Zhou, 1996) and xenophobia (Mora & Dávila, 2005) as factors that may make this a negative indication of integration.   We did not perform a full scan of the literature, but given that this project is not theoretical in nature, we thought it best to leave this data point value-neutral.

  Another important issue for this data point is that the sample sizes became very small when we cut it by immigration status as well as race and ethnicity – and as a result we could not report on all categories.   If a data point has fewer than 30 cases, we considered it too unstable. This is a lower threshold than we usually use, but too much data was rendered unstable, otherwise, and variation by race/ethnicity is high.

- **Industry**: The reported percentage represents the share of civilian, working aged (25-64 years old), employed people attached to the field in question. For example, in San Diego, 24% of immigrants are in Professional and Related Services. This universe is the same for industry data reported on page two of each regional insert.

- **LPRs and Voting Population**: The voting-eligible population includes all U.S. citizens age 18 and over. The estimates of adult LPRs eligible for naturalization are based on the data obtained from the Office of Immigration Statistics (see the "Civic Infrastructure for Naturalization" section) and include all LPRs who attained status between 1985 and 2005, but had not naturalized as of 2010.

- **Overskilled Immigrant Workers**: More information on this can be found earlier, in the explanation of indicators used for the Economic Snapshot category.

- **Language Abilities**: In the back page table display "Top Languages Spoken in Immigrant Households" and the graph "Immigrant English Skills by Recency of Arrival," only immigrants ages 5 and older are included.

References

Batalova, J., & Fix, M. (2008). *Uneven Progress: The Employment Pathways of Skilled*

    Immigrants in the United States. Washington, D.C.: Migration Policy Institute. Retrieved

    from http://www.migrationpolicy.org/pubs/BrainWasteOct08.pdf

Cataldo, S. (2008). *Immigration in the Rockies*. The 2008 Colorado College State of the Rockies

    Report Card. Colorado Springs, CO: Colorado College State.

Children Now. (2004). *California Report Card: Focus on Children in Immigrant Families*. Oakland,

    CA: Children Now.

Fix, M., Laglagaron, L., Terrazas, A. M., & McHugh, M. (2007). *An Analytic Framework for*

    Developing an Immigrant Integration Strategy for Los Angeles County. Washington D.C.:

    Migration Policy Institute.

Fix, M., McHugh, M., Terrazas, A. M., & Laglagaron, L. (2008). *Los Angeles on the Leading Edge:*

    Immigrant Integration Indicators and their Policy Implications. Washington, D.C.: The

    Migration Policy Institute.

Jenson, J. (1998). *Mapping Social Cohesion: The State of Canadian Research* (No. F03).

    Canadian Policy research Networks Inc. Retrieved from

    http://www.rwbsocialplanners.com.au/spt2006/Social%20Cohesion/Mapping%20social

    %20cohesion.pdf

Kahn, C., Kingsley, T., & Taylor, G. (2010). *Shared Indicators System Framework (Draft)*.

    Washington, D.C.: National Neighborhood Indicators Partnership (NNIP). Retrieved from

    http://www.neighborhoodindicators.org/sites/default/files/publications/indicframewor

    k100810.pdf

Kaufmann, D., & Kraay, A. (2008). Governance Indicators: Where are we, where should we be going? *The World Bank Research Observer*, *23*(1). Retrieved from mpra.ub.uni-muenchen.de/8212

Marcelli, E. A., & Lowell, B. L. (2005). Transnational Twist: Pecuniary Remittances and the Socioeconomic Integration of Authorized and Unauthorized Mexican Immigrants in Los Angeles County. *International Migration Review*, *39*(1), 69–102. doi:10.1111/j.1747-7379.2005.tb00256.x

Mora, M. T., & Dávila, A. (2005). Ethnic group size, linguistic isolation, and immigrant entrepreneurship in the USA. *Entrepreneurship & Regional Development*, *17*(5), 389–404. doi:10.1080/08985620500275612

Pastor, M., & Ortiz, R. (2009). *Immigrant Integration in Los Angeles: Strategic Directions for Funders*. Los Angeles, CA: USC Center for the Study of Immigrant Integration. Retrieved from http://dornsife.usc.edu/pere/documents/immigrant_integration.pdf

Pitkin, J., & Myers, D. (2011). A Summary Period Measure of Immigrant Advancement in the U.S. *Demographic Research*, *24*, 257–292.

Portes, A., & Zhou, M. (1996). Self-Employment and the Earnings of Immigrants. *American Sociological Review*, *61*(2), 219–230. doi:10.2307/2096332

Rytina, N. (2011). *Estimates of the Legal Permanent Resident Population in 2010* (Population Estimates) (p. 4). U.S. Department of Homeland Security; Office of Immigration Statistics.

Seghetti, L. M., Vina, S. R., & Ester, K. (2006). *Enforcing Immigration Law: The Role of State and Local Law Enforcement* (CRS Report for Congress No. RL32270). Washington, D.C.:

Congressional Research Service and ilw.com. Retrieved from

http://www.ilw.com/immigrationdaily/news/2006,0912-crs.pdf

Spoonley, P., Peace, R., Butcher, A., & O'Neill, D. (2005). Social Cohesion: A Policy and Indicator

Framework for Assessing Immigrant and Host Outcomes. *Social Policy Journal of New*

Zealand, (24), 85–110.

The Drum Major Institute for Public Policy. (2009). *Suffolk County: Legislative Scorecard*. Port

Washington, N.Y.: Long Island Wins. Retrieved from

http://www.longislandwins.com/downloads/20100802-Scorcard_FINAL.pdf

Vigdor, J. L. (2008). *Measuring Immigrant Assimilation in the United States.* (Civic Report No.

53). New York, NY: Manhattan Institute for Policy Research. Retrieved from

http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED

501689

Wood, G. (2000). *Concepts and Themes: Landscaping Social Development* (SD Scope Paper No.

9). Social Development Department, Centre for Development Studies, University of

Bath. Retrieved from http://staff.bath.ac.uk/hssgdw/lansocdev.pdf

ACKNOWLEDGEMENTS

P: 213.821.1325 | F: 213.740.5680
E: CSII@USC.EDU | HTTP://CSII.USC.EDU

950 W. JEFFERSON BLVD, JEF 102
LOS ANGELES, CA 90089-1291

University of Southern California