Evaluating College Readiness for English Language Learners and Hispanic and Asian Students

A paper presented at the annual meeting of the American Educational Research Association
Vancouver, Canada

Min Wang
Keyu Chen
Catherine Welch

April 2012
Abstract

Group differences in growth trajectories are of interest when a common growth model and developmental scale are used for all students. Parallelism of growth trajectories found in this study provides validity evidence for measuring growth and readiness of not only culturally but also linguistically diverse groups using the Iowa Assessments.
Evaluating College Readiness for English Language Learners and Hispanic and Asian Students

Objectives

In a global economy, countries are competing for top talent and resources in order to stay ahead. For each individual, higher education or professional training is necessary to succeed in this competitive world. According to a U.S. Bureau of Labor Statistics report, nearly half of all new jobs through 2018 will require at least some form of postsecondary education or training (U.S. Bureau of Labor Statistics, 2009). However, scores from 2011 ACT tests indicated that only about 25% of recent U.S. high school graduates meet College Readiness Benchmarks (CRBs) (ACT, 2011).

Since the announcement of the Race to the Top program in 2009, college and career readiness (CCR) has become another measure of quality of K-12 education. Various studies have been conducted to define, implement and assess it. Among them, some research has focused on the CCR information of diverse populations (Bustamante, Slate, Edmonson, Combs, Moore, & Onwuegbuzie, 2010). To better understand the performance and college readiness in culturally and linguistically diverse groups, long-term analyses are needed to monitor the growth trends of these students in reaching CCR targets. In order to address this issue, the present study aims to establish validity evidence in using the developmentally scaled Iowa Assessments for measuring growth and CCR of English language learners (ELLs), Hispanic, and Asian students. We propose to:

1. Examine the average performance of groups of interest for coincident patterns of growth along a vertical scale;
2. Assess the percentages of students meeting CCR targets in different groups;
3. Explore the achievement gaps between diverse groups and all students based on CCR targets in Reading, Language, Mathematics, and Science.

**Theoretical Framework**

In recent years, the enrollment of students who are ELLs, Hispanic and Asian have been increasing. From the statistics provided by U.S. Department of Education, among the 49.9 million students who were enrolled in public school system in the 2007-2008 academic year, 10.7 percent were ELLs (Aud, Hussar, Planty, Snyder, etc., 2010), 22% were Hispanic, and 3.7% were Asian (National Policy institute, 2010). In the state of Iowa, the percentage of ELL students in the 2009-2010 academic year more than doubled from that reported 10 years earlier. The percentage of Hispanic students increased from 2.6% in the 1997-1998 academic year to 8% in the 2009-2010 academic year, and from 2.1% to 5.9% for Asian students (Iowa Department of Education, 2010).

Under such circumstances, a better understanding of group results on growth and CCR is indispensable to policy makers, educators, and test developers. Both ACT (2009, 2010a, and 2011) and SAT (College Board, 2011) have periodically reported national and state statistics on the college and career readiness for the graduating class. However, the information was categorized by racial or ethnic groups (i.e., African American Black, Caucasian American White, Hispanic, and Asian American), while information about ELLs as a separated group is scarce. Measuring growth is important in evaluating ELLs because many ELLs may have the same ability as their peers, but might not perform as well as their peers due to lack of sufficient academic English proficiency and cultural background knowledge (Miller-Whitehead, 2005). Supporting this, the Test Standards have emphasized the effect of language proficiency in uses and interpretations of test scores and encourage test developers to collect validity evidence.
concerning whether scores differ in meanings for test takers with linguistically diverse backgrounds from scores of the population of all examinees as a whole (AERA, APA, NCME, 1999). In this study, we provide validity evidence for measuring the growth and readiness of not only culturally but also linguistically diverse groups using the Iowa Assessments.

Of specific interest in this study is the developmental pattern (a trajectory) of average student scores relative to progress toward a college readiness benchmark. Group differences in these trajectories would call into question the use of a common growth model and developmental scale for all students. Similarity on parallelism of growth trajectories would support the validity of scale scores for making “on track” interpretations of growth relative to college readiness, regardless of student background and educational experiences.

**Design and Methods**

**Vertical Scale**

To assess students’ progress and growth, the use of tests with vertical scales is necessary. A vertical scale is developed when special assessments that are appropriate for students of different developmental levels are administered across grades. Performance on each of the test levels is related to a single numerical scale that reflects the growth patterns of students (Kolen & Brennan, 2004). The Iowa Standard Score Growth Model used in the Iowa Assessments related to this study is such a metric and is capable of tracking students’ growth over years.

**College and Career Readiness Track for the Iowa Assessments**

ACT’s College Readiness Benchmarks (CRBs) were empirically derived and relate to the minimum scores needed for students to have a high probability to success in the first-year credit-bearing college courses, such as English Composition, College Algebra, and Biology (ACT, 2010b). Because career readiness demands the same level of knowledge and skills as college
readiness (ACT, 2008), CRBs could be used to gauge the readiness of students on both academic setting and working world.

In a previous study, researchers had identified scale scores (SS) on the Iowa Assessments that correspond to the ACT established CRBs for the 11th grade (Furgol, Fina & Welch, 2011), and comparable SSs for earlier grade levels were then derived through the properties of the vertical scale. Therefore, a College and Career Readiness (CCR) track is able to be created from grade 6-11 for the Iowa Assessments. The underlining assumption is that, if students consistently score above the identified SS for a specific content area on the corresponding earlier grade level test, they will be above the grade 11 cut score associated with the predicted college readiness scale score. As a result, by looking at the SSs for a student, one can judge whether this student is on the CCR track. See the above figure for an example of how SSs can be used to monitor growth and progress toward a college readiness target (Welch & Dunbar, 2011).

**Instruments**

In this study, scores on subtests of the Iowa Assessments (grade 6-8 on the Iowa Tests of Basics Skills and grade 9-11 on the Iowa Tests of Educational Development) were used to conduct the analyses. The subtests are Reading, Language, Mathematics and Science. These subtests were chosen because the tests for grade 11 have similar content specifications as the ACT, and were used to carry out a linking study to establish CCR targets for the Iowa Assessments (Furgol, Fina & Welch, 2011).
The Iowa Assessments represent samples of items on an achievement continuum that measure student growth from kindergarten to twelfth grade in core academic areas important for success in college (Welch & Dunbar, 2011). The most recent forms of the assessments have been carefully aligned to the Common Core State Standards (CCSS) and The Iowa Core, and are “developed in collaboration with teachers, school administrators, and experts, to provide a clear and consistent framework to prepare our children for college and the workforce” (Common Core State Standards, 2010).

**Samples**

The dataset used in this study is a matched cohort of students enrolled in public and private schools in Iowa in the 2002-2003 school year, who have a grade 6 test score in at least one of four subjects of interest and who may or may not have grade 7-11 test scores in subsequent school years until 2007-2008, provided the enrollment status. Depending whether a student took the *ACT* test or not, this cohort was partitioned into two subsets: “*ACT* takers” and “non-*ACT* takers”. See Table 1 presented below for a demographic breakdown of the sample. Because complete census data were not available for the year of 2006-2007, the scale score for grade 10 was calculated using the average of scale score of grade 9 and 11.

**Table 1**

*Total Numbers for Students with or without *ACT* Scores in the 2002-2003 Academic Year*

<table>
<thead>
<tr>
<th></th>
<th>All students</th>
<th>Non-<em>ACT</em> takers</th>
<th><em>ACT</em> takers</th>
<th>Percentage of <em>ACT</em> takers</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELLs</td>
<td>449</td>
<td>307</td>
<td>142</td>
<td>31.63</td>
</tr>
<tr>
<td>Hispanic</td>
<td>937</td>
<td>615</td>
<td>322</td>
<td>34.36</td>
</tr>
<tr>
<td>Asian</td>
<td>415</td>
<td>137</td>
<td>278</td>
<td>66.99</td>
</tr>
<tr>
<td>Other*</td>
<td>25886</td>
<td>11549</td>
<td>14337</td>
<td>55.38</td>
</tr>
<tr>
<td>Total</td>
<td>27687</td>
<td>12608</td>
<td>15079</td>
<td>54.46</td>
</tr>
</tbody>
</table>

* Students identified as other racial, demographical, or linguistically backgrounds.
Design and Data Analysis

Descriptive statistics were calculated so that the research questions could be addressed. In addition, specific analyses are described that focus on similarities of growth trajectories and differences between proportions of students “on track” for CCR.

**Split-plot MANOVA.**

In order to evaluate parallelism of average performance growth curves and group differences among growth curves, a series of $2 \times 6$ split-plot multivariate analyses of variance (MANOVAs) were conducted, one for each subgroup of interest from one sub-sample for one subject area (e.g., ELLs from ACT takers in math, or Asian students from non-ACT takers in Reading, etc.). The within-subjects factor was the repeated measure of the Iowa Assessments scale score for one group of students from one subset in one subject area over six years (e.g., scale score for ELLs from subset of ACT takers in Mathematics grade 6-11, or scale score for Asian students from subset of non-ACT takers in Reading grade 6-11, etc.). The between-subjects factor was the subgroup students of interest versus the remaining students in the sample, such as ELLs compared to remainder of ACT takers. In order to establish independent groups for each comparison, the remaining students in each subset were used as a reference group for all the statistical analysis, instead of using group of all students. Patterns in growth trajectories for demographic groups were calculated with effect sizes based on standard multivariate tests (e.g. Johnson & Wichern, 2008). Only observations with complete data on all variables were included in the calculation of effect sizes for parallel slopes and achievement gaps.

**Standardized proportion differences.**

To investigate the college readiness status and gaps between groups, firstly, proportions of students reaching CCR targets were examined for the above thirty-six combinations. Then
taking the complement to a given focal group as a reference, standardized proportion differences between each subgroup and the reference group were calculated using pooled within-groups standard error estimates.

**Results**

The correlations between the four subject areas of the grade 11 Iowa Assessments and the corresponding *ACT* tests were calculated for ELLs, Hispanic and Asian students. The modest to strong relationship found in previous study (Furgol, Fina & Welch, 2011) holds for grades 5-10. On average, the correlations are most close to all students for Hispanic and Asian students, but are about 0.1 lower for the ELLs. These results are reported in Table 2.

**Table 2**

*Average Correlation between the Iowa Assessments and ACT over Six Years*

<table>
<thead>
<tr>
<th></th>
<th>Reading</th>
<th>Language</th>
<th>Mathematics</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELLs</td>
<td>0.67</td>
<td>0.65</td>
<td>0.65</td>
<td>0.54</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.72</td>
<td>0.74</td>
<td>0.70</td>
<td>0.61</td>
</tr>
<tr>
<td>Asian</td>
<td>0.77</td>
<td>0.76</td>
<td>0.77</td>
<td>0.64</td>
</tr>
<tr>
<td><strong>All students</strong></td>
<td><strong>0.73</strong></td>
<td><strong>0.74</strong></td>
<td><strong>0.74</strong></td>
<td><strong>0.62</strong></td>
</tr>
</tbody>
</table>

**Average Performance Trends**

Similar growth trend between subgroups and all students can be clearly seen in Figures 1-4 in Appendix A. Occasional small interactions between groups and average SSs were observed in certain subjects, subgroups or datasets. To summarize the patterns in all of the Figures given in Appendix A, a multivariate effect size was calculated. This statistic range between 0 and 1.0 and takes on values close to 0 when the hypothesis of parallel profiles is supported by the data. The results are reported in Table 3. The consistently trivial effect sizes in Table 3 for departure from parallel trajectories indicate that the developmental trends of ELLs, Hispanic and Asian
students are very similar to the trends of all students. There is little, if any, evidence in these effect sizes to suggest group differences in growth trajectories.

Table 3

*Effect Size* for Departure from Parallel Trajectories

<table>
<thead>
<tr>
<th>Subject</th>
<th>Dataset</th>
<th>ELLs</th>
<th>Hispanic</th>
<th>Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>All</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>ACT takers</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Non-ACT takers</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Language</td>
<td>All</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>ACT takers</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Non-ACT takers</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Mathematics</td>
<td>All</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>ACT takers</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Non-ACT takers</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Science</td>
<td>All</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>ACT takers</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Non-ACT takers</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

*Effect Size* = 1 - Wilks' Lambda (Johnson & Wichern, 2007)

To better illustrate the results in Table 3, Figure 1 presents the average performance for subgroups in Reading as well as for all students as an example of the parallel trajectories observed. Although from the graph the profile for Asian students overlaps the profile of all students, the effect size of 0.01 even for this comparison indicates no practical significance. Hence, parallel growth trajectories seem to be true for the three datasets, four subject areas, and three demographic groups of interest in this study.

In addition, the achievement gaps observed in the graphs were investigated as well. The standardized mean difference between one subgroup and the reference group were tabulated and are presented in Table 4.
Figure 1

Growth Trends for ELLs, Hispanic and Asian Students in Reading in All Students Dataset.

Table 4

<table>
<thead>
<tr>
<th>Subject</th>
<th>Dataset</th>
<th>ELLs</th>
<th>Hispanic</th>
<th>Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>All</td>
<td>0.35</td>
<td>0.23</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>ACT takers</td>
<td>0.37</td>
<td>0.21</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Non-ACT takers</td>
<td>0.24</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>Language</td>
<td>All</td>
<td>0.32</td>
<td>0.24</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>ACT takers</td>
<td>0.30</td>
<td>0.21</td>
<td>-0.05</td>
</tr>
<tr>
<td></td>
<td>Non-ACT takers</td>
<td>0.22</td>
<td>0.16</td>
<td>-0.04</td>
</tr>
<tr>
<td>Mathematics</td>
<td>All</td>
<td>0.35</td>
<td>0.27</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>ACT takers</td>
<td>0.36</td>
<td>0.24</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Non-ACT takers</td>
<td>0.25</td>
<td>0.20</td>
<td>0.09</td>
</tr>
<tr>
<td>Science</td>
<td>All</td>
<td>0.35</td>
<td>0.26</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>ACT takers</td>
<td>0.37</td>
<td>0.24</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Non-ACT takers</td>
<td>0.26</td>
<td>0.18</td>
<td>0.14</td>
</tr>
</tbody>
</table>

*Effect Size = Cohen's d, positive effect size favors the reference group*
With respect to standardized performance differences between diverse groups and all students, no effect was observed for Asian students, which means that Asian students performed as well as all students. An effect size around 0.2 for Hispanic students indicates that there is an achievement gap between Hispanics and the reference group, but the gap is smaller than the gap between ELLs and the reference group. One noticeable pattern is that, a relatively smaller effect size was observed in the subset of non-ACT takers, when the effect sizes were compared across datasets. This disparity is clearly illustrated in Figure 2 below, that the gaps between subgroups and all students are smaller for the dataset with only non-ACT takers, and are very similar for the dataset with all students and the dataset with only ACT takers. Furthermore, the abovementioned gaps between the subgroups and all students stay similar across all subjects. On average, ACT takers performed about 14 Iowa scale score points better than all students and about 32 Iowa scale scores better than non-ACT takers across all subjects and over the six years included in this study.

Figure 2

Average Performance Profile for ACT Takers and non-ACT Takers in Reading
Thus, in terms of improvement of growth over time, our first finding that the parallelism of growth trajectories between special population and their peers, confirms efforts made by educators and parents to promote learning for all students. It indicates that all students are actually obtaining growth at the same rate under well-developed educational plans. However, special interventions are needed to help these diverse learners grow at a faster rate so that they could be as ready as their peers for future education and careers.

**Percentage on CCR Track**

The parallelism pattern found above also provides strong evidence on the equitability of applying the CCR track from the Iowa Assessments to all students including the diverse groups of students considered in this study (Furgol, Fina & Welch, 2011). Comparisons between different groups of students were conducted with respect to the percentage of students on the CCR track. Overall, 31% of the *ACT* takers, 20% of all students, and 7% of the non-*ACT* takers are on track base on their scores from the Iowa Assessments, which is defined as being on track for all four subjects (*ACT*, 2010). Notable in all the datasets, the percentage of Asian students that is on track in all four subject areas is the highest among subgroups of interest, with about 34%, 27% and 13% for *ACT* takers, all students, and non-*ACT* takers, respectively. Nevertheless, ELLs and Hispanic students are much farther behind their classmates. Since the continuous population increases of Hispanic students and ELLs, imagine what the CCR status graphs would look like had they been as ready as their peers. Graphs of the percentages described above are given in Figure 3 and in Appendix B.
Figure 3

*Percentage of Students Meeting CCR Targets across Datasets*

Although the developmental trends for the diverse groups are very similar in subject areas across datasets and over years, the percentages of students reaching CCR targets vary by groups of interest, subjects and datasets.

To further examine the percentages of students reaching CCR targets, standardized proportion differences were calculated. These are similar to student’s t statistics and are presented to provide a better statistical metric for evaluating achievement gaps. In terms of standardized proportion differences between subgroups and the reference group, the percentage differences between subgroup and all students are very small for Asian students, but are markedly larger for Hispanic students and ELLs. As for subject areas, the smallest proportion differences observed are in Reading for non-\textit{ACT} takers, while the largest observed difference is in Science for all subgroups across all datasets. Graphs displaying these statistics are given in Figure 4.
Figure 4
Standardized Proportion Differences on CCR Track between Subgroups and Others
In terms of CCR proportion gaps, as illustrated in Figure 4 above, Asian students and the reference group were most close. The percentage of Asian students reaching CCR targets is more than that of reference group in Reading, but is much less in Science for non-ACT takers. For Asian students who are ACT takers, the standardized proportion difference with the reference group is positive in Language for both ACT and non-ACT takers, which indicates that more Asian students reach the targets than students in the reference group. However, as previously mentioned, the reference group had fewer than half its students above the CCR targets.

For Hispanic students and ELLs, the standardized proportion differences for reaching CCR targets are quite similar among ACT takers for every subject area. But for non-ACT takers, the readiness gaps vary across subjects. The largest standardized proportion difference for Hispanic students shows up in Science, which is similar for Asian students, and the smallest is in Reading in non-ACT taker group. For ELLs, the differences are consistent and are all negative across datasets. Interestingly, the biggest and the smallest gaps observed for ELLs were both in Reading, and were for ACT takers and non-ACT takers, respectively. This phenomenon indicates that the predicted CCR condition of ELLs who did not take ACT is much closer to their peers than ELLs who took ACT on perhaps the most language-dependent test — Reading.

**Significance**

The preparation for college and career readiness is a continuous process throughout elementary and secondary education, which requires long-term monitoring and a high quality assessment system. Different models may be developed to specify how students’ progress and how the CCR status to be assessed. However, the model presented here may provide some ideas of how an assessment system can use a vertical scale, strong content alignment to the Common Core Standards, statistical linkage with an admission test and longitudinal data to provide
validity evidence on growth and in applying the CCR track to all students, including these special populations, regardless of college aspirations.

With this long-term monitoring information, parents and educators would determine that whether a group is on track to college and career readiness, begin reacting early to help younger students achieve their goals, and help teachers provide support strategies and intervention at the right moment. Moreover, it helps educators in identifying the strengths and weaknesses of different groups, which would be helpful for educational administrators in improving outcomes for their students, especially for students from minority groups, and in closing achievement gaps among students. As a result, we would be closer to the goal of identifying and preparing every student for college and career by 2020 (U.S. Department of Education, 2010).
References


Figure 1

*Average Performance Trend for All Students*
Figure 2

*Average Performance Trend for ACT Takers*
Figure 3

*Average Performance Trend for non-ACT Takers*
Figure 1

Percentage on CCR Track for All Four Subject Areas
Figure 2

Percentage on CCR Track for All Students
Figure 3

Percentage on CCR Track for ACT Takers
Figure 4

Percentage on CCR Track for non-ACT Takers

Appendix B