



Selecting One Story and Hiding Others: How AYP Chooses the Portrayal of a School

Peter Clyde Martin

Ithaca College

Citation

Martin, P. C. (2011). Selecting One Story and Hiding Others: How AYP Chooses the Portrayal of a School. *Current Issues in Education*, 14(1). Retrieved from <http://cie.asu.edu/>

Abstract

This article discusses how the Adequate Yearly Progress (AYP) accountability mechanism of No Child Left Behind makes use of supposedly objective standardized test scores to describe schools in a certain way when the same results could serve to draw very different conclusions.

Examining the proficiency scores of students from a specific middle school that is determined to be unequivocally failing according to AYP, the article shows how the same data could be used to characterize this school in ways that are more complex and emphasize its possible accomplishments. In addition to disaggregating test scores according to AYP-prescribed student categories, the study accounts for length of time students have been enrolled in the school and

considers those students who do not present characteristics that NCLB views as putting them at risk of school failure as a separate category of their own. The study shows that the portrait of the school that is drawn by including these additional categories is quite different from that presented by the existing AYP reports. It is suggested that, in addition to questions regarding the validity of high-stakes assessments for the evaluation of student learning, the very way in which the data is presented in AYP reporting makes tacit choices that are disturbing, need to be highlighted, and whose implications ought to be considered.

Keywords: School accountability, urban education, education reform, equity, special education, English language learners

About the Author(s)

Author: Peter Clyde Martin

Affiliation: Ithaca College

Address: Department of Education, 194 Phillips Hall, 953 Danby Road, Ithaca, NY 14850

Email: pmartin@ithaca.edu

Biographical information: Dr. Martin is on the faculty of the Education Department at Ithaca College. He holds a doctoral degree in Bilingual Special Education from The George Washington University. His research interests focus on the areas of teacher collaboration, vision-based and transformational schooling, and serving the needs of English language learners considered to be at risk of educational failure.



Current Issues in Education

Mary Lou Fulton Teachers College • Arizona State University
PO Box 37100, Phoenix, AZ 85069, USA

Selecting One Story and Hiding Others: How AYP Chooses the Portrayal of a School

When the No Child Left Behind Act of 2001 (NCLB) was voted into law with overwhelming support from a wide range of political currents and interest groups, its promise was to finally ensure that in the United States high academic achievement would be attained by all students regardless of race, background, income, or other special circumstances. For the first time, all public schools were required to implement annual student testing for a systematic measurement of educational effectiveness and were held accountable for all students being academically proficient by 2014. NCLB insisted that schools be held accountable for the success of all students, including those who are poor, are learning English as a second language, or have a diagnosed disability.

Issues of Accountability

The scope and historic nature of NCLB's accountability system has been lauded on many fronts. It has been pointed out that for the first time all the nation's public schools are subject to clearly defined educational goals for student achievement (The Education Trust, 2004) and that there is now a consistent system for determining school success (Hall, Wiener, & Carey, 2003; Yell, Katsiyanna, & Shiner, 2010). Credit has been given to how, as a consequence, schools that historically failed to adequately educate their students are no longer allowed to simply keep doing so (Hall et al., 2003). Advocates for students with disabilities have remarked that NCLB's insistence that schools be held accountable for the academic proficiency of all students has led to greater inclusion of special education students in grade-level instruction (Hardman & Dawson, 2008).

It can be argued, however, that the enforcement arm of NCLB, namely the Adequate Yearly Progress (AYP) accountability mechanism with its underlying high-stakes assessment

tools, has also ended up serving as a way to conceal individual student differences (and thereby US social inequities). Ironically, AYP may have become a means to gloss over disparities when it could instead be a tool to highlight differences and potential connections between student characteristics and academic achievement and thus truly hold schools accountable.

This article is an examination of one middle school in Washington, DC where the AYP process has been used to draw one particular portrait of the school—tell one particular story—by presenting one view of test scores and the student population while ignoring other views and characteristics of the very same scores and student body. By proclaiming one conclusive—supposedly objective—view of the school, the AYP process as it is implemented in fact reflects a particular choice of what to tell and what not to tell, choosing one story over others.

Case of Washington, DC: AYP and the AYP Report

The particulars of how a school educates its student body, with the characteristics that make it unique, could be expected to dictate how it is portrayed. Given their history of perceived failure, it would seem that public schools in Washington, DC would be especially keen on having portrayal mechanisms in place that highlight not only what they are achieving, but also the difficulties they contend with.

NCLB was in fact designed to highlight those achievement gaps and educational failures that schools and districts were not noticing, divulging, or effectively addressing. In itself, this initiative was nothing new. The Improving America's School Act of 1994 (IASA) had already reacted to findings of persistent lags and gaps in student achievement by requiring that states implement academic standards and assessments. This, it was thought, would keep schools and states from simply ignoring problems and provide data to guide improvement measures. The effectiveness of these initiatives, however, was undermined by a lack of consistency across states

and the absence of any comprehensive accountability system (Yell et al., 2006). NCLB, as the continuation of IASA, sought to rectify this by setting a clear objective for all schools across states (100% proficiency in math and reading by 2014) and backing it up with an accountability system based on AYP reporting, the details for which would be defined on a state level. By insisting on rigorous assessments of all students nation-wide and holding schools and districts accountable for results, the federal government seemed to take away the option of hiding or ignoring educational failure.

In order to meet federal AYP requirements of assessing at least 95% of all students in grades 3 through 8 and once in high school in Reading and Mathematics, the D.C. Office of the State Superintendent of Education (OSSE) has mandated that the D.C. Comprehensive Assessment System (DCCAS) test be given each spring. Standard scores are calculated and categorized as Below Basic, Basic, Proficient, and Advanced based on single-score cut off points. In accordance with NCLB, all but very few students with disabilities are required to take the test in its standard form. Scores for English language learners (ELL) are not counted for AYP purposes during students' first year in the United States. Both ELLs and students with disabilities are allowed specific accommodations for taking the test. These accommodations are considered to have no effect on the construct or the standardized nature of the test.

In Washington, DC every school is subsequently issued a one-page "AYP Report" that lists the percentage of students who were assessed and of those who scored Proficient or above. Also listed are the percentages for subcategories by ethnicity, special education classification, ELL classification, and students who are economically disadvantaged. Finally, the report indicates whether, for each student category, the school reached the predetermined state-wide proficiency target. Indeed, since NCLB mandates that 100% of all students in every category

score Proficient by 2014, states have set annually increasing targets in order to meet the federal requirement. Thus in 2007 schools in DC were required to meet proficiency targets of 43.58% in Reading and 40.54% in Math in order to make AYP. In 2008 the target jumped to 57.69% in Reading and 55.41% in Math. Based on whether the target is reached in the given year, the official report states whether the school made AYP or whether it is in “need of improvement.” Depending on the number of consecutive years the school has not made AYP, a continuum of remedial measures are mandated, with the eventual threat of reconstitution. Schools that do not make AYP but show at least a 10% decrease in students who did not score Proficient are given a ‘Safe Harbor’ designation that freezes their AYP status until the following year. The summary AYP Report is disseminated to the public and published in the local newspaper without further analysis of the student population, of the actual standard scores, or of any measure of statistical validity. Essentially, this cursory glance at some of the assessment data establishes an official, public view of a school as, for example, successful, struggling to become successful, or incontestably failing. In addition to formal state-mandated repercussions, this AYP report also serves to guide students, families, teachers, and administrators in their choice of what schools to enroll or work in. While other states have a dual system that allows for AYP to be down-played in the public view in favor of a more variable state accountability determination, this has not been the case in DC. The significance of this single-page AYP Report and its consequences for the individual schools can therefore not be overstated.

While on the surface the very succinctness and quantitative nature of the data on the AYP report would seem to go a long way toward establishing clear-cut, measurable goals and holding schools accountable in an undisputable way, the narrative the report tells is in fact neither clear-cut nor undisputable. By disaggregating the assessment data in one way and not in others, and

by identifying some categories and not others, a specific story is selected among many. What becomes in effect the official story of the school is selected rather than evident as the only one possible.

Due to low proficiency rates common in DC public schools, the degree of failure as publicized in the AYP report is important as well. Indeed, for the school year 2007-2008 only 39% of all secondary school students scored proficient in Reading and 36% in Math. Since these results are 19% and 20%, respectively, below what was stipulated, many schools are arguably not only under pressure to make AYP, but also—if they do not—to not fall significantly below the local average (DC Government, 2007). The AYP report is thus critical beyond whether a school has made AYP. For stakeholders the difference between being a school that struggles to try to make AYP and one that is irrevocably failing could be significant.

In the case of a middle school that has to accept for at most three years any student coming from any elementary school with any language and economic background and a possible disability, the AYP report is equivalent to its public story. This story, in turn, is not dictated by the specifics of the school, the demographic makeup or mobility of its population, or even the quality of its feeder schools.

Assessing Student Differences Rather than Student Learning

There is evidence that scores on standardized tests closely correlate with demographic factors rather than with educational achievement directly resulting from instruction. In their examination of student information in testing data from multiple states, Marchant, Paulson, & Shunk (2006) found that characteristics of the test-takers accounted for up to seventy percent of the variance among scores in a same state. Similarly, Welner (2005) shows that if two students

are given the same interventions to help increase their achievement, the outcomes will be different for each based on factors specific to the learner.

A number of authors have argued that AYP measures do not, in fact, account for why one student's scores are lower than another's (Welner, 2005; Hershberg, 2008; Koretz, 2008; Jennings & Corcoran, 2009). Koretz (2008) suggests that achievement on AYP assessments "reflects the cumulative effect of many factors, both educational and not," thus hardly making it an objective measure of instructional quality. Welner (2005) argues that test scores are, if anything, the reflection of the effect of a large number of different and varied factors and stakeholders. According to him, "the truth is that each—school and student—bears some responsibility, along with the state, the school district, the family, the community, peer groups, libraries, and various other people and institutions including the federal government. And the truth is that if it were possible to measure the actual contributions of each to student test scores, we would find varying proportions for each community, family, student, teacher, and school. A more rational NCLB would acknowledge both those truths."

What are we (not) evaluating? An inevitable consequence of the ambiguous relationship between instruction and test scores is the AYP system's failure to do what it so adamantly aims to ensure, namely, a way to fairly and clearly evaluate schools and thus hold them accountable for student learning (Barton, 2006). According to Hershberg (2008), data indicates that there is greater variance in student achievement within schools than among schools, thus highlighting the flaws in how test scores are used for evaluation purposes. Koretz (2008) points out that once student background factors are controlled for, there is not much variation among scores. There therefore seems to be little direct connection between school quality alone and scores on AYP measures.

Who are we (not) targeting? An additional flaw in the use of AYP measures to evaluate educational effectiveness relates to how many schools try to increase the percentage of their students who score proficient. In their study of increases in 5th grade test scores in Chicago Public Schools, Neal & Schwanzenbach (2007) found that the introduction of test-based accountability measures did in fact correlate with an increase in scores for students in the middle of the achievement distribution. There was no evidence, however, of gains among lower students and only mixed evidence of increases among the higher students. Thus, while schools demonstrated progress in student achievement, these concerned only a certain group. According to the authors, “schools face weak incentives to devote extra attention to students who are either already proficient or who have little chance of becoming proficient in the near term” (Neal & Schwanzenbach, 2007). In a study of the effect of the test-based accountability system and teaching practices in Texas, Booher-Jennings (2005) found evidence of educational triage whereby teachers diverted resources to those students believed to be close to passing the assessment bar and to those whose scores would most affect the school’s accountability rating. Ironically, AYP provides clear incentives to focus instructionally on some students only and thereby neglect others.

Economically disadvantaged students. The percentage of economically disadvantaged students in schools impacts AYP determinations. Such inequity is confirmed by data on the achievement of students in certain sub-groups. There is ample evidence that schools, both regular and charter, with high percentages of students who qualify for free and reduced lunch receive lower average scores than do schools with wealthier students (Escamilla Mahon, Riley-Bernal, & Rutledge 2003; Jennings & Corcoran, 2009).

English language learners. ELLs and the schools that serve them also receive generally

lower test scores, based on the fact that they speak English as a second language and not because they receive an inferior education. While there are strong arguments that including ELLs in AYP-related assessments and accountability measures ought to improve the quality of the education they receive by leading teachers to establish higher expectations, the consequence is often a purposeful educational exclusion (Menken, 2008). According to No Child Left Behind, making AYP requires that ELLs who have been enrolled in US schools for at least one year be administered the same assessments as their native-English speaking peers—in English—even though these tests were developed for native speakers of English (Menken, 2009). As a consequence, serious questions have been raised as to the actual validity of administering these assessments to ELLs—let alone holding anyone accountable based on the results (Government Accounting Office, 2006; Menken, 2008).

Not surprisingly, then, students with limited proficiency in English score generally lower than native speakers (Abedi & Dieta, 2004; Government Accounting Office 2006; Abedi, 2009). Ultimately, any content-focused achievement test, when administered to ELLs, will in effect function as a test of English proficiency (Menken, 2008, 2009). English proficiency, however, is not what is supposed to be measured when evaluating whether a school is making adequate yearly progress. The assessments, in this case, do not test what they claim and the accountability system uses them to make judgments it cannot fairly make. Consequently, schools with large numbers of ELLs are, in effect, unfairly punished for serving a particular population (Escamilla et al., 2003; Ga'ndara & Baca, 2008). As Menken (2008) states in her study on discriminatory testing practices in New York public schools, “language has become a liability in our quest for accountability.”

Students with disabilities. Similarly, there is strong evidence that students with disabilities have far lower proficiency results than the law mandates. Typically, students' academic levels are determined to be at least two years below their grade if they qualify as having a disability. Nevertheless, close to all of them are administered the same one-size-fits-all test as students who do not have disabilities. Generally there is a significant performance gap between special education and general education students (Abedi, 2009). Indeed, it is common for schools not to make AYP because of the test performance of their special education subgroup (Eckes & Swando, 2009; Cole, 2006). It should be noted, however, that a number of advocates for the education of students with disabilities have lauded their inclusion in the assessment process as a way to ensure—as NCLB intends—that schools are held fully accountable for their learning (Hardman & Dawson, 2008).

Not surprisingly given the heterogeneity of types and degrees of disability and given the nature of many prevalent processing disorders, there is evidence of a wide range, inconsistency, and unpredictability of group scores that makes the validity of evaluating schools with students with disabilities using AYP measures highly questionable. For example, in one large-scale multi-state study that examined progress of the special education subgroup for individual schools over time, both gains and drops were so significant from one year to another that it became difficult to make inferences about performance trends for students with special needs as a group (Thurlow, Quenemoen, Altman, & Cuthbert, 2008).

Given that AYP-related tests were not designed for many students with disabilities, the low test scores and their unreliability should not come as a surprise. In a study of scores from California, Texas, and Florida, the low scores of students with disabilities was attributed largely to the fact that these constituted the only AYP subgroup whose defining characteristics involved

actual limitations on the ability to learn. It was concluded that differences in proficiency levels between students with and without disabilities were evident from the onset and that these remained over time, regardless of the progress made by each subgroup. The scores of students with disabilities stayed lower than those of any other category. AYP requirements therefore effectively mandate that special education students make progress faster than their peers who do not have disabilities (Eckes & Swando, 2009). In fact, the very principle of NCLB and AYP that all students are to reach a common level of proficiency contradicts the Individuals with Disabilities Education Act's (IDEA) mandate that students are held to expectations that are individualized based on their needs (Albritte, Mainzer, & Ziegler 2004). Furthermore, there is little evidence that assessment accommodations that special education students are entitled to are indeed effective in making the tests more accessible (Bowen & Rude, 2006). Ultimately, this makes it more difficult for a school that has significant numbers of students with disabilities to make AYP than one that does not. Again, student population is the determiner of a school's "quality" according to AYP rather than the level of instruction (Eckes & Swando, 2009).

English language learners with disabilities. If ELLs and students with disabilities are inherently less likely to score Proficient on assessments, it follows that students who are both ELL and qualify for special education would be most severely handicapped in the context of AYP (Minnema, Thurlow, Van Getson, & Jimenez, 2006; Barrera, Shyyan, Liu, & Thurlow 2008; Abedi, 2009).

In a study comparing test scores of ELLs, non-ELLs with disabilities, ELLs with disabilities, and students who were in none of these categories, there was evidence of a clear performance gap favoring students who were not ELLs and did not have disabilities. Significantly, the gap was highest with students who were ELL and had disabilities. They had

the lowest scores as a group (Abedi, 2009).

It is suggested that these students' failure to score Proficient is a given because of the nature of the test. Abedi (2009) argues that "due to the impact of linguistic and cultural factors and also because of their disabilities, assessments that are developed for mainstream students may not be providing valid outcomes for these students." Given disabilities that, for the most part, have to be impacting their learning in order to be diagnosed as such, and given the fact that they are in the process of learning the language of the test, ELLs with disabilities are, as a group, almost guaranteed to do poorly and their schools to be judged as doing a poor job teaching them.

The Case of the Bilingual Community Academy

Deciding whether to take into account poverty, disability, and language background in the way the AYP gavel falls when judging the quality of a school is critical in determining the public profile of any school that struggles not to lose potential students to others that are better-regarded. The public story that the AYP system tells, emphasizing some aspects of the data and not others, is puzzling given the problems in validity and significance of standardized test scores, but also risks misrepresenting the struggles and successes of the schools it describes. This article focuses on the hidden story of one such school, deemed to be unequivocally and miserably failing according to one view of its assessment scores, while a more detailed scrutiny of these same, AYP-targeted assessment-results tell a story that is still problematic, yet far more complex, nuanced, and arguably more hopeful.

The Bilingual Community Academy ("ABC" for its Spanish name, Academia Bilingue de la Comunidad) was a middle school in Washington, D.C. whose appearance was heavily marked by its AYP report card that made public its very low test scores and status as failing according to NCLB. Established as a charter school in 2005 with a Spanish-English bilingual

program, it attracted a small student population that was almost exclusively low-income, mostly came from low-performing public schools, and was almost evenly split between Hispanic and African-American students. Depending on the year, special education students made up between 21.65% and 28.85% of the population. While ABC served grades 6 through 8, and did not have a feeder school, students entered in all three grades. It was thus never clearly established as a three-year program, with students coming and going between one school year and the next.

Table 1 gives a demographic breakdown of the student population over the first three years of the school's existence (2005 through 2008).

Table 1

ABC Student Population by Year and Subgroup*

Group	Year	Number	Percentage of Total
African-American	2006	26	49.1
	2007	53	55.2
	2008	56	53.8
Hispanics	2006	27	50.1
	2007	43	44.8
	2008	48	46.2
Caucasian	2006	0	0
	2007	0	0
	2008	0	0
Females	2006	25	47.2
	2007	55	57.3
	2008	54	56.2
Males	2006	28	52.8
	2007	41	42.7
	2008	50	43.8
Special Education	2006	13	24.5
	2007	21	21.9
	2008	30	31.2
Non-Special Education	2006	40	75.6
	2007	75	78.1
	2008	74	69.8
LEP/NEP**	2006	23	43.4
	2007	41	42.7
	2008	50	48.1
Non-LEP/NEP	2006	30	56.6
	2007	55	57.3
	2008	57	51.1
Economically Disadvantaged	2006	53	100
	2007	92	958
	2008	104	100
Non-Economically Disadvantaged	2006	0	0
	2007	0	0
	2008	0	0
Total	2006	53	100
	2007	96	100
	2008	104	100

*These represent the number of students enrolled, which may be slightly different from the number of students assessed on the DCCAS.

**This constitutes a subgroup among students categorized as English language learners according to district regulations.

The official story: A failing school with failing scores. Based on the one-page AYP Report that every school in Washington, DC is issued upon disaggregation of the previous year's assessment, ABC consistently failed as a school. Table 2 shows the summary portion of the AYP report for all students. ABC never came close to reaching the benchmarks in any of the achievement-related indicators. The school never reached the Safe Harbor mark that is available for schools that do not reach the absolute benchmarks necessary to make AYP, but still achieve substantial growth in their proficiency scores. Based on the official story told by the report and spread by newspapers ABC was nothing more than yet another very bad school in a city where bad schools seemed to be the norm.

Table 2

Percent Proficient on the 2008 DC-CAS*

Subject Tested	Number Tested	Number Proficient/Adv.	Percent Proficient	Made AYP?
Reading	98	23	23.71%	No
Math	101	16	16.0%	No

* Each of the three grades tested was subject to the same proficiency target and received the same AYP designation.

Looking more closely: The hidden stories. However, there are untold stories underneath this official one that are contained in details that AYP omits. As the literature explains (Welner, 2005; Hershberg, 2008; Koretz, 2008; Jennings & Corcoran, 2009), there are questions as to the significance of high-stakes tests for evaluating the achievement of different

sub-populations. Indeed, if the assessment results—and therefore their significance—are affected by issues of poverty, language proficiency, and disability, then it would follow that they mean something quite different depending on the students' circumstances. Scores should be viewed with the understanding that there are doubts as to their validity for a number of subpopulations and that, in effect, they are to be considered most meaningful when evaluating students who are neither ELLs nor individuals with disabilities.

ABC's actual student population. Examining ABC's test results while taking into consideration both the reality of its particular student population and some of the above-mentioned doubts as to the appropriateness and validity of the test for certain subgroups, a fairly complex picture emerges that suggests the importance of broadening the ways AYP results are reported. While, as the data will show, ABC was a struggling school that would not have made AYP even if the scores had been reported differently, its degree of failure and apparent hopelessness was associated primarily with demographic and turnover factors that the report does not highlight. Given that the AYP report does not provide confidence intervals and that the objective of this article is not to estimate a parameter but simply to show how the same interpretive rules for the AYP report could be used to tell a more nuanced narrative, confidence intervals will not be used here.

Several authors have reported on the negative impact of poverty on test scores (Escamilla et al., 2003; Jennings & Corcoran, 2009). As the near entirety of ABC's student body came from low-income homes (depending on the year, between 95.8% and 100% of the students qualified for free and reduced lunch), it is unfortunately not possible to provide comparative data in this regard.

One important point of comparison when attempting to make sense of what the scores actually represent and that is not reported on for AYP is the amount of time students spent in the school. Indeed, one might assume that a school's effectiveness would be more definitely demonstrated by students who have been enrolled for some time than for students who are new. This appears to be corroborated by the data.

Indeed, there was a noticeable difference in scores between students who had newly transitioned to the school and others who had two or three years of continuity at ABC. If one measured the proficiency status of returning students only the percentage scoring Proficient was higher (30.9% in Reading, or 7.19% more than the total, and 21.8% in Math, or 5.8% more than the total, see Table 3). Conversely, the status of students who were in their first year at ABC was lower (14.29% in Reading, or 9.42% less than the total and 8.8% in Math, or 13% less than the total). While scores remain far below the AYP target, one of the untold stories is therefore how newness to the program may have had a negative impact on the likelihood of scoring Proficient. Given the large number of new students (43.9% in Reading and 45.5% in Math), any difference here is worthy of attention.

Table 3

Percent Proficient—Returning Students

	Total Students	New Students*			Returning Students		
Subject Tested	Percent Proficient	Number Tested	Number Proficient	Percent Proficient	Number Tested	Number Proficient	Percent Proficient
Reading	23.71%	43	6	14.29%	55	17	30.9%
Math	16.0%	46	4	8.8%	55	12	21.8%

* Only includes students enrolled since the beginning of the school year

This trend of better test results over time is confirmed when we separate scores of students who have attended the school for more than one year (see Table 4). Indeed, the percentage of students scoring Proficient increases significantly with each year they remain at ABC. While the reasons for this student mobility are beyond the scope of this study, turnover (42.11% and 61.05%, depending on the year) can be negatively linked to the assessment scores, a fact that is omitted from the official story of the school.

If we look at students who spent all three years in the school—as the design of ABC’s program called for—the difference in scores is moderate in Math, but becomes remarkable in Reading. In Reading we note a jump from 33.98% below the AYP target for all students (and 43.4% for new students) to 18.8% below the AYP target for students who were in their third year. If in Reading they had been considered an AYP subgroup (which they were not) in order to hold the school accountable for academic achievement over time, then theirs would be a story of a school struggling to narrow the gap rather than one of utter academic failure.

Table 4

Percent Proficient—Returning Students after 2 and 3 years

% Proficient		Total Students	1 st year students	2 nd year students	3 rd year students
Reading	Number tested	98	42	38	18
	Number proficient	23	6	10	7
	Percent proficient	23.71%	14.29%	26.32%	38.89%
Math	Number tested	101	45	38	18
	Number proficient	16	4	8	4
	Percent proficient	16.0%	8.89%	17.78%	22.22%

The large percentage of students with disabilities should also be underlined as a significant aspect of the school and its results. As mentioned in the literature (Eckes & Swando, 2009; Cole, 2006), there are questions as to the validity of using these test scores and benchmarks to evaluate the learning of students with IEPs. Not surprisingly, then, average scores are very different when general education and special education students are counted separately. This is particularly significant in the context of telling the story of ABC, given the

high percentage of special education students who took the high-stakes tests (32.7% of the total in Reading and 31.7% in Math).

The percentage of students scoring Proficient jumps considerably if results for general education students are viewed separately. Indeed, while none of the students with disabilities scored Proficient in either subject, 35.29% of General Education students made Proficient in Reading (or 11.58% more than the total) and 25.71% made Proficient in Math (or 9.71% more than the total) (See Table 5). Scores for general education students as a group are not mentioned in the AYP report.

Table 5

General Education Students

Subject	Total Students	Students with Disabilities			Students without Disabilities		
	Percent Tested Proficient	Number Tested	Number Proficient	Percent Proficient	Number Tested	Number Proficient	Percent Proficient
Reading	23.71%	32	0	0%	66	23	34.85%
Math	16.0%	32	0	0%	69	16	23.18%

Research has pointed to a lack of validity of high-stakes tests to measure the academic achievement of ELLs, questioning whether they don't in fact assess English language proficiency instead (Menken, 2008). Given that ELLs constituted close to half the population of the school during each of the three years, it seems appropriate to identify two separate language-related stories for ABC.

In both Reading and Math the percentage of ELLs who scored Proficient was lower than for native speakers. ELLs scored 7.35% lower in Reading than non-ELLs and between 6.99% lower in Math. For ELLs, of course, the question ultimately becomes how much of a story these results actually tell other than that they are in fact in the process of acquiring the English language.

Table 6

Non-English Language Learners & English Language Learners

Subject	Non-English Language Learners			English Language Learners		
	Number Tested	Number Proficient	Percent Proficient	Number Tested	Number Proficient	Percent Proficient
Reading	52	14	26.92%	46	9	19.57%
Math	52	10	19.23%	49	6	12.24%

So far we have considered how students with characteristics that undermine the validity of the test results have done worse on AYP measures than others who did not have that particular characteristic. In many ways, this investigation follows the logic of AYP which also examines results for the total population and those particular subgroups that are seen to be at a disadvantage. We have shown how in the case of ABC these disadvantages do in fact seem to be substantiated by the test scores, but also how the large size of these subpopulations plays a part in determining the AYP results for the school’s population as a whole, thus indicating that the narrative embedded in these test results is far more complex and nuanced than the AYP report indicates.

While the report accounts for the results of those students considered to be disadvantaged, it is important to note that it does not do so for those students that the test appears to actually be designed for. Indeed, at ABC a minority of the students were not English language learners and did not have a documented disability. If the AYP report had a separate category for general education students who are native speakers of English, they would, as a group, tell a different story. In this unlabeled subgroup 48.28% of the students scored Proficient in Reading (or 23.28% more than in the total population and 9.41% less than the AYP target) and 34.48% in Math (or 19.1% more than in the total and 20.93% less than the AYP target, see Table 7). It is also to be noted that, had the AYP target for Reading not increased from the previous year, these students would, as a subgroup, have made AYP. This story is ignored in the official AYP report, substituting instead a single narrative about an irrevocably failing school.

Table 7

Percent Proficient— General Education: Non-English Language Learners

Subject	Total Students	General Education-ELLs			General Education-Non-ELLs		
	Percent Tested Proficient	Number Tested	Number Proficient	Percent Proficient	Number Tested	Number Proficient	Percent Proficient
Reading	23.71%	36	9	25.00%	29	14	48.28%
Math	16.0%	39	6	15.38%	29	10	34.48%

Telling a Different Story

In addition to determining the need for remedial or even punitive measures, AYP reports play an important role in describing for the general public the work accomplished by an individual school. This function of the AYP mechanism in turn needs to be reconciled with the much-documented fact that schools that are more diverse economically and/or racially are less likely to make AYP than schools that are not (Kim & Sunderman, 2005; Balfanz, Legters, West, & Weber, 2007). In their study of high schools, Balfanz et al. (2007) found that schools that have numerous student subgroups (denoting diversity) rarely make AYP. Kim and Sunderman (2005) propose that requiring all subgroups to meet the same proficiency goals places schools with high levels of poverty and diversity at a disadvantage. The authors suggest that African-American and Hispanic students, in particular, often belong to other AYP subgroups as well and that schools with large minority populations are therefore even less likely to make AYP (Kim & Sunderman, 2005). It seems important that AYP clearly present and underline these factors as relevant to the educational gaps of a school rather than gloss over them.

Supposedly, the AYP process, as established under NCLB, provides the irrefutable story of a school's success or failure to educate its students. Supported by what is supposed to be the transparency and evidence of numbers derived from standardized test results, the narrative provided by AYP is to be an objective, factual report based on the sense that 'numbers don't lie.' According to the AYP formula, the story of the ABC middle school is simply one of complete educational failure. Low percentages of students scored Proficient among the total population, economically disadvantaged students, ELLs, and students with disabilities, in both Reading and Math.

Looking further into the data used for AYP purposes, we find that the clarity and simplicity of this story hinges on the fact that certain subgroups are singled out and others are not, and that the effects of specific student characteristics on proficiency results are not being taken into account at all. Currently, NCLB mandates that AYP be determined and reported for the following four subgroups of students:

- Economically disadvantaged students
- Students with disabilities
- English language learners
- Students from major racial and ethnic groups

In order to use the AYP system to truly render a plausible portrait of the school, the findings discussed here suggest additional categories. First of all, it is urged that the complementary categories for those listed above be used as well and given equal weight so that they, too, become full parts of the story. Here the following categories are suggested:

- Students who are not economically disadvantaged
- Students without disabilities
- Students who are not English language learners
- Students who are not economically disadvantaged or in special education or English language learners.

If one purpose of AYP is to make explicit and force schools to address educational gaps among subgroups, not listing results for this second set of categories inevitably leads to an incomplete and potentially false picture.

Furthermore, the findings of the article demonstrate a need to reflect the transient nature of part of our student population and the fact that schools' accountability for the achievement of

students who have been enrolled for several years will undoubtedly be different than for students who have not. The following two categories are therefore suggested, at a minimum:

- Students who have attended the school for no more than one year
- Students who have attended the school for more than one year

Once we include in our story the widely-documented truths that being in the process of learning the language one is tested in, having a disability that is defined based on one's low academic achievement, and being new to a school that is still working with one's prior learning are all factors that have a negative impact on test scores, the story changes. Instead, the story becomes that of a school that is actively struggling to achieve a degree of success while serving a largely poor and diverse population of students, many of whom have personal characteristics that put them at a disadvantage when it comes to meeting the requirements of AYP. Rather than present an 'objective truth' thanks to its underlying quantitative data, the AYP process reflects a failure to tell the story of the population of a given school, thus hiding those very differences that it was supposed to highlight and not reporting on outcomes directly linked to learning rather than demographics.

References

- Abedi (2009). English Language Learners with Disabilities: Classification, Assessment, and Accommodation Issue. *Journal of Applied Testing Technology*, 10(2).
- Albritten, D., Mainzer, R., & Ziegler, D. (2004). Will students with disabilities be scapegoats for school failures? *Teaching Exceptional Children*, 36 (3), 74-75.
- Balfanz, R., Legters, N., West, T. C., & Weber, L. M. (2007). Are NCLB's measures, incentives, and improvement strategies the right ones for the nation's low-performing schools? *American Educational Research Journal*, 44(3), 559-594.
- Barrera, M., Shyyan, V., Liu, K. K., & Thurlow, M. L. (2008). *Reading, mathematics, and science instructional strategies for English language learners with disabilities: Insights from educators nationwide* (ELLs with Disabilities Report 19). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Barton, P. (2006). *"Failing or "succeeding" schools: How can we tell?* Washington, DC: American Federation of Teachers.
- Booher-Jennings, J. (2005). Below the bubble: 'Educational triage' and the Texas accountability system. *American Educational Research Journal*, 42(2), 231-268
- Bowen, S. K. & Rude, H. A. (2006). Assessment and students with disabilities: Issues and challenges with educational reform. *Rural Special Education Quarterly*, 25(3), 24-30.
- Cole, C. (2006). *Closing the achievement gap series: Part 3. What is the impact of NCLB on the inclusion of students with disabilities?* Bloomington, IN: Center for Evaluation and Education Policy, Indiana University.

DC Government (2007, August 17). District of Columbia releases school achievement data.

Washington, DC: Author. Retrieved from <http://newsroom.dc.gov/show.aspx/agency/seo/section/2/release/11684/year/2007>

Eckes, S. E. & Swando, J. (2009). Special Education subgroups under NCLB: Issues to consider. *Teachers College Record* 11 (11), 2479-2504.

Escamilla, K., Mahon, E., Riley-Bernal, H., and Rutledge, D. (2003). High-stakes testing, Latinos, and English language learners: Lessons from Colorado. *Bilingual Research Journal*, 27(1), 26-50.

Government Accountability Office. (2006). *No child left behind act: Assistance from education could help states better measure progress of students with limited English proficiency*.

Washington, DC: Author.

Hall, D., Wiener, R. & Carey, K. (2003). What new "AYP" information tells us about schools, states, and public education. *Washington, D.C.: The Education Trust*.

Hardman, M. L. & Dawson, S. (2008). The impact of federal public policy on curriculum and instruction for students with disabilities in the general classroom. *Preventing School Failure*, 52(2), 5-12.

Hershberg, T. (2005). Value-Added Assessment and Systemic Reform: A Response to the Challenge of Human Capital Development. *Phi Delta Kappan*, 87(4).

Jennings, L. J. & Corcoran, S. P. (2009). "Beware of Geeks Bearing Formulas": Reflections on Growth Models for School Accountability. *Phi Delta Kappan*, 90(9), 635-639.

Kim, J. S. & Sunderman, G. L. (2000). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, 34(8), 3-13.

Koretz, D. (2008). A Measured Approach. *American Educator*, Fall 2008.

- Marchant, G. J., Paulson, S. E., & Shunk, A. (2006). Relationships between high-stakes testing policies and student achievement after controlling for demographic factors in aggregated data. *Education Policy Analysis Archives*, 14(30).
- Menken, K. (2008). *English learners left behind: Standardized testing as language policy*. Clevedon, Avon: Multilingual Matters.
- Menken, K. (2009). Policy failures: No Child Left Behind and English language learners. In S. Groenke & A. Hatch (Eds.), *Critical pedagogy and teacher education in the neoliberal era: Small openings*. Berlin, Germany: Springer.
- Minnema, J.E., Thurlow, M.L., Van Getson, G.R., Jimenez, R. (2006). *Including English language learners with disabilities in large-scale assessments: A case study of linguistically-diverse populations* (ELLs with Disabilities Report 14). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Neal, D.& Schwanzenbach, D. W. (2007). *Left Behind by Design: Proficiency Counts and Test-Based Accountability*. (NBER Working Paper Nr.13293). Cambridge, Mass.: National Bureau of Economic Research, 2007.
- Sparks, S. D. (2010). Study finds fewer ‘dropout factory’ schools. *Education Week*, 30(14),12 - 13.
- The Education Trust (2004). The ABCs of “AYP.” Raising achievement for all students. *Washington, DC: Author*.
- Thurlow, M., Quenemoen, R., Altman, J., Cuthbert, M. (2008). *Trends in the participation and performance of students with disabilities* (Technical Report 50). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Yell, M. L., Katsiyanna, A. & Shiner, J. G. (2006). The No Child Left Behind Act, Adequate

Yearly Progress, and students with disabilities. *Teaching Exceptional Children*, 38(4), 32-39.

Welner, K. G. (2005). Can irrational become unconstitutional? NCLB's 100% presuppositions. *Equity & Excellence in Education*, 38, 171-179.



Current Issues in Education



<http://cie.asu.edu>

Volume 14, Number 1

ISSN 1099-839X

Authors hold the copyright to articles published in *Current Issues in Education*. Requests to reprint CIE articles in other journals should be addressed to the author. Reprints should credit CIE as the original publisher and include the URL of the CIE publication. Permission is hereby granted to copy any article, provided CIE is credited and copies are not sold.



Editorial Team

Executive Editors

Lori Ellingford
Andrew Darian

Layout Editor
Jennifer Wojtulewicz

Faculty Advisers
Gustavo E. Fischman
Jeanne M. Powers
Debby Zambo

Assistant Executive Editor
Krista Adams

Copy Editor
Lucinda Watson

Section Editors
Hillary Andrelchik
Meg Burke
Douglas Deiss
Miriam Emran
Tracy Geiger
Sarah Heaslip
Melinda Hollis
Afzal Hossain

Seong Hee Kim
Yoonsu Kim
Alaya Kuntz
Angeles Maldonado
Carol Masser
John Michael
William Mitchell
Elizabeth Reyes
Lindsay Richerson

Rory Schmitt
Tapati Sen
Jennifer Shea
Kara Sujansky
Melissa Tarango
Andrew Tesoro
Jill Wendt
