

Teachers' Construction of Expectations and Adaptation of Practice in Response to Ability

Grouping, Labels and Performance Data

Christopher Harrison

The University of North Carolina at Chapel Hill

This research was conducted with funding from the Institute of Education Sciences (R305C10023). The opinions expressed in this article are those of the author and do not necessarily represent the views of the sponsor.

Introduction

It has been common practice in schools, historically, to categorize students according to their academic “potential” or perceived performance (Slavin, 1987; Oakes & Lipton, 1992). These practices manifest in a number of ways - from broad systems of explicit academic tracking, to teachers’ individual efforts to differentiate instruction according to the needs of students in the classroom environment. Educators undertake these efforts, ostensibly, with the goal of increasing educational productivity; significant evidence indicates, however, that these efforts are founded upon, and continually reify, outdated norms and beliefs regarding the nature of intelligence, learning and differences between groups of students. Further, research indicates that educators may act on these beliefs by altering their practice and framing their expectations for student performance in ways that perpetuate existing social inequities and limit students’ educational outcomes.

Today, questions surrounding the issue of ability grouping are complicated by the rapid expansion of educators’ access to rich, comprehensive student performance data. These data are often framed as providing school actors with objective information regarding students’ ability and learning growth over time. Additionally, strong school accountability structures reinforce educators’ use of such data to diagnose students’ academic needs, deficiencies and potential. Ultimately, educators are expected translate their perceptions of these data into pivotal decisions regarding students’ educational experiences – choices that include grouping students across and within classrooms, as well as modifications to the curriculum and instruction experienced by groups of students.

In order to more fully understand the ways that new sources of student performance data may be interacting with and, potentially, reinforcing school cultures shaped by decades of ability grouping and tracking in schools, I draw from evidence collected during a comparative case study of high schools in Florida. In doing so, I first construct a brief review of how grouping practices – including tracking – have developed over time, and how current trends in data use may be shaping such practices now. I then discuss the nature of the data from which I derive my findings, and the analytic process used to explore them. Finally, I present my findings, followed by a brief discussion of their implications.

Tracking and Detracking: A Brief Review

The practice of grouping students according to their ability has a long history in the American public school system. One of the most contested practices – systemic tracking of students based on their perceived ability or academic performance – has been a topic of significant study and debate in the educational community. Slavin (1987), in his synthesis of the research on academic tracking, notes that the theory of action underlying such structures centers on enhancing academic productivity. Through grouping students according to ability, he notes, educators may be able to more ably divert resources and tailor instructional practice to students' specific learning needs.

Some research (Gamoran, 1992; Gamoran & Mare, 1989; Lee & Bryk, 1988; Slavin 1987) provides limited and conditional evidence that certain manifestations of academic tracking may serve the purpose of aligning educational experiences to students' individual needs. Gamoran (1992), for example, defines several characteristics of tracking systems that may mediate student outcomes under such structures – including the relative mobility afforded to

students in the grouping structure, and the “inclusiveness” of performance groups. He finds that systems that allow for relative inclusiveness of the “higher” academic track, along with mobility between academic groups, may allow for broad improvement in student achievement. Lee & Bryk (1988), similarly, find that the greater mobility afforded to students in Catholic schools, and greater inclusiveness of the “academic” track, led to higher levels of “advanced” course taking and post-secondary continuation than in the public system.

Detractors of systemic efforts to group students on the basis of ability – particularly rigid structures of academic tracking – assert, however, that such systems often entrench and exacerbate existing social inequities (Oakes, 1992; Oakes, et al., 1990; Oakes & Guiton, 1995; Wells & Oakes, 1996). Research indicates that students’ placement into academic tracks, for example, may be more closely related to their racial or socio-economic backgrounds than their academic needs or performance. Further, significant evidence (Gamoran & Mare, 1989; Hoffer, 1992; Natriello, Pallas & Alexander, 1989) indicates that schools – in response to pre-conceived notions regarding student ability or pressure from external agents like parents – may concentrate effort and resources with students in “higher” or tracks, surrounding “lower-track” students in cultures of low expectation and perceived deficiency.

Evidence that systems of academic tracking contribute significantly to inequities in educational outcome and opportunity resulted in a strong push from numerous policy actors to dismantle structures of systemic ability grouping (Oakes & Lipton, 1992). Research (Argys, Rees & Brewer, 1996; Burris, Heubert & Levin, 2006; Yonezawa, Wells & Serna, 2002) indicates that attempts to construct more heterogeneous learning environments paid dividends for at-risk students, particularly when coupled with broader, systemic reforms. Several scholars (Lucas & Berends, 2002; Oakes & Lipton, 2002; Oakes, Wells, Jones & Datnow, 1997),

however, assert that pernicious structural and cultural aspects of academic “tracking” remain in public schools, even in the wake of technical reform.

The Persistence of Ability Grouping Practices and Structures in Schools

Despite significant pressure to “de-track” public schools, evidence indicates that ability grouping – and the potentially negative outcomes of such practices – persists across schooling organizations. A number of studies (Clotfelter, Ladd & Vigdor, 2005; Congar, 2005; Lucas & Berends, 2002), for example, indicate that the structural framework underlying systems of academic tracking appears to be resilient. Strong relationships continue to exist between students’ racial or socio-economic backgrounds and the courses that they take – particularly in high schools (Lucas & Berends, 2002). Minority students, and students from distressed socio-economic backgrounds, are more likely than their peers to be placed in remedial courses (Clotfelter, Ladd & Vigdor, 2005). Research also finds that these students – who are often most in need of educational resources – are more likely to be segregated into lower-performing schools and to be taught by inexperienced or novice teachers (Clotfelter, Ladd & Vigdor, 2005; Congar, 2005).

Further, several studies (Oakes & Lipton, 1992; Wells & Oakes, 1996; Oakes, Wells, Jones & Datnow, 1997; Yonezawa, Wells & Serna, 2002) report the persistence of school cultures borne of the principles underlying ability grouping structures; these “cultures” are grounded in deeply rooted beliefs about learning and development – including the fixed nature of intelligence and differences in academic potential between groups of students. Within these cultures, educators form powerful beliefs about students that inform their expectations and, ultimately, their practice. Decades of research (Brophy & Good, 1970; Goldring, et al., 2009;

Good, 1987; Hallinger & Murphy, 1986; Klem & Connell, 2004) documents the impact of variable expectations for student performance on student outcomes; this evidence indicates that groups of students immersed in cultures of low expectation experience worse outcomes than students challenged by rigorous expectations. Research (Gamoran, et al., 1995) also indicates that teachers operating within such cultures translate their expectations, reified by structures and practices of ability grouping, into real differences in instruction.

The Growing Role of Performance Data

Increasingly, schools emerging from this historical context – many of which carry with them school cultures, norms and beliefs shaped by ability grouping practices like systemic tracking – are gaining access to, and making use of, rich sources of student performance data (Cohen-Vogel & Harrison, in press; Firestone & Gonzalez, 2007; Guskey 2003; Halverson, Grigg, Prichett & Thomas, 2007; Ingram, Louis & Schroeder, 2004; Louis et al., 2010). These performance data include a variety of student performance indicators, from teachers' own diagnostic efforts to externally-derived indicators like scores on state standardized tests. The latter, in particular, are lent significant validity and importance within state accountability frameworks - especially under the auspices of policy initiatives like No Child Left Behind and Race to the Top (Anderson, Leithwood & Strauss, 2010; Firestone & Gonzalez, 2007).

Research also supports the notion that educators across states and districts are responding to the signals accountability systems send about such data, aligning their decision-making processes and practices accordingly. Administrators report, for example, that they engage in systematic efforts to structure school learning environments in response to what performance data tell them about their students; principals, for instance, may engage in “staffing to the test”,

assigning students in accordance to the prior performance and targeting interventions for specific students (Cohen-Vogel, 2011; Cohen-Vogel & Harrison, in press; Louis, et al., 2010) . Teachers, similarly, report that they are under heavy pressure to utilize systemic performance data to engage in decisions regarding curricular pacing and instruction (Cohen-Vogel & Harrison, in press).

Based on this research, schools appear to be under pressure from a number of potentially inimical forces – specifically, from dueling imperatives to “de-track” and dismantling structures and cultures of ability grouping, and the urgency of responding to accountability frameworks that legitimize performance data as a means for knowing students’ potential and which press for greater educational productivity. Schools that exhibit a greater ability to drive learning gains in traditionally underserved students – minority students, students from disadvantaged socio-economic backgrounds, and English language learners – may navigate these tensions very differently than schools that have been less successful in closing performance gaps.

Methods

In order to understand whether and why these differences between “higher” and “lower” performing schools may exist, this study utilizes data gathered the National Center for Scaling Up Effective Schools (NCSU) – a federally funded research center aimed at understanding why certain schools appear to be particularly able to promote growth in traditionally underserved sub-groups, and scaling their effective practices within and across school organizations.

Data for this paper, and the NCSU’s broader work, are drawn from a comparative case study of large and demographically diverse comprehensive high schools (Rutledge, Cohen-Vogel & Osborne-Lampkin, 2012). Working in Florida, we began by using a simple value-added

achievement model (VAM) to estimate the relative performance of all of the state’s high schools (Sass, 2012). The estimated fixed effect for each high school in the state was put in rank order and classified by deciles of value-added. These analyses identified only one Florida district with multiple high and low-performing schools serving traditionally underperforming students; we selected this district for our work.

A variety of data were collected from four case study schools in this district – two of which (Beacon Hills and Pine Coast) were identified as “higher performing” according to our VA analysis, and two of which (Cypress Cove and Bay Mountain) were “lower” performing - during three week-long visits in the 2010-2011 school year. A demographic description of each of our partner schools is provided in Table 1, below. This study relies upon a sub-sample of the data collected during this process – specifically, interviews conducted with four assistant principals and twelve 10th-grade teachers of ELA, Math and Science in each of the four case study schools. Interview protocols were designed to elicit participants’ perceptions regarding a variety of topics, including their use of performance data, the ways in which adults and students were assigned to classes and the nature of instructional practice in the school.

Table 1. *Demographic Profile of Case Study Schools: 2010-2011*

School	Enrollment	% FRPL	% MIN	% ELL
Cypress Cove	2100-2300	45-55	55-65	5-10
Bay Mountain	1800-2000	60-70	55-65	10-15
Beacon Hills	2200-2400	45-55	65-75	5-10
Pine Coast	2800-3000	30-40	50-60	5-10
District Average	2327	52	73	9

Note: Shaded cells indicate schools identified as higher value-added using our value-added model; non-shaded are lower value-added. To ensure confidentiality, school names are pseudonyms, and demographic values have been converted to ranges. Demographic data were derived from the School Public Accountability Reports (SPAR) compiled by the Florida Department of Education. FRPL = Free and Reduced Price Lunch; MIN = Minority (African American, Hispanic, Asian & Other); ELL = English

Language Learners. District averages were calculated with data from all non-alternative, non-charter public high schools in the district (N=29).

Following the data collection process, participants' responses were systematically analyzed through directed content analysis (Patton, 2002). I started by analyzing the data categorically, first assigning basic, descriptive codes drawn from a review of the literature on tracking, "de-tracking" and school data use. I constructed codes, for example, to identify *educators' construction of performance groups*. To understand the ways in which educators respond to the knowledge they derive from structures and cultures of ability grouping, I constructed codes for *educators' expressed expectations* (high or low) and *educators' alteration of practice*. Finally, in order to better understand the varying ways that teachers might navigate the tension between educational productivity and equity, I constructed codes identifying educators' *intended outcomes* (productivity or equity).

Further allowing themes to emerge from the data inductively (Miles and Huberman, 1994), I then constructed "sub-categories" under *educators' construction of performance groups*, comprised of *broad data-based performance groups* and *FCAT categories*. Basic codes identifying *teachers' alteration of practice* were similarly expanded, to include secondary codes identifying participant responses describing *targeting of academic support*, *targeting of social support*, *differentiated instruction*, *modification of pacing*, *modification of rigor*, *homogeneous grouping* and *heterogeneous grouping*. Throughout the coding process, I engaged in systematic effort to seek out and identify disconfirming evidence – what Corbin & Strauss (2008) term "cases that don't fit the pattern" (p. 298). In the case that such evidence was found, it was used to refine the above coding framework.

Findings

Educators' Construction of Performance Groups

Across all four schools, teachers and administrators reported that the framing of categories for understanding students' academic potential – both structurally and through individual practice – was a general norm. Further, they reported that performance data, including scores on the state's standardized achievement test (FCAT) and other benchmark tests administered by the district, served as a foundation for doing so. One teacher captured this phenomenon, sharing that “we are not supposed to label kids but we are being told to label kids. You can't segregate kids into parts of the rooms, but we are being told to segregate. It's the ultimate Catch-22”.

Across three of our case study schools – Cypress Cove, Bay Mountain and Pine Coast – numerous administrators and teachers expressed that data informed their perception that students were separable into broad “performance groups” – one particularly common frame of reference was the “bottom 30%”. One teacher shared, for instance, that:

right now, we are having a meeting about how to help the lower 30 percent move up [...] Like first we had to identify, we had a meeting and you had to identify the 30 and what it is [...] do you know how to identify the lower 30, here are some strategies you might use with them.

While the “bottom 30%” was, by far, the most commonly reported performance grouping, participants also constructed other explicit “labels” to reference student ability – including the “lowest quartile” and “higher performing” or “lower performing” students. In most cases, these broad groups were founded upon educators' interaction with student performance data.

In many cases, educators – particularly teachers- “drilled down” from broad performance groups and engaged in “labeling” students according to their FCAT scores. One teacher, for example, argued that one of the key challenges in his school “is [that] they like to place level ones and twos in places where they don't belong. Like my honors class, 75 percent are level one

and two. I have gifted kids in there, and the gifted kids they are at a disadvantage because they have to deal with the ones and twos. They are bored, so they want to shout out obscenities, or do nothing because they don't care". These references often drew stark lines that separated "ones", "twos" and "threes" – students either below or just meeting proficiency according to the FCAT – from the higher performing "fours" and "fives".

Beacon Hills stood out from the other schools in that several participants – generally teachers - expressed that they came to "know" their students through "breaking down" data to understand students' individual "strengths" and "weaknesses". In some cases, this conception of "strengths" and "weaknesses" was paired with language centered on FCAT levels – again, however, such language was framed around understanding students' individual performance across particular skills. One teacher, for example, shared that

I will look at their reading levels to make sure they are on level, and if they are like level one in reading that concerns me. So, if I see a student that's going to be a struggling reader, our Virtual Counselor breaks them up into strands so I can see if they have a issue with the main idea, with reference and research, with vocabulary, things like that. So that helps. It's really easy to use because if they are struggling below a three, it's in red.

Students' position within the perceptual hierarchy framed by many educators across the case study schools – as a member of the "bottom 30%", or as "level one" or "level two" FCAT performer – was complicated by and, often, clashed with the systemic placement of students across classrooms. This was, largely, endemic to all four schools and resulted from conflict between local and district practice. Students, according to educators across the district, were placed through the use of a data-based "assignment matrix". One teacher explained the system, reporting that "the big thing is the reading score. They can't take an honors course unless they have three or above in reading [...] I think that's across the board for math and English whether they are in honors or regular class".

One difference between the higher-performing and lower-performing schools, however, was the relative willingness on the part of school actors to accept the results of the data-based student assignment system. Actors across both lower-performing schools expressed a strong sentiment that the district's propensity for increasing the heterogeneity of classes – specifically, placing “lower level” students in “higher level” courses - was a negative practice, and clashed with their perceptions of which students were honors or AP “material”. A teacher in Cypress Cove, for example, asserted that

I think the only thing going wrong with every school is that they keep integrating the abilities when there is nothing wrong with streaming students into classes that they can be successful at. Because the higher ones can bring up the lower ones, but then the higher ones suffer and come down, and we are forgetting about the high fliers, that we really need for the future, for our engineers, and doctors, and lawyers, and things like that.

In Bay Mountain, several participants went further, sharing that there was a strong movement in the school toward creating a wholly separate space for the school's “high performers”. One administrator framed this effort as an attempt to boost the school's population of “four's and five's, kids that would help the school tremendously in terms of competition, kids in the valedictorian, salutatorian [range]”; in doing so, he shared, the school hoped to “raise the bar. It may separate them more, and I don't want to create an elitist group on campus, but if we can create competition, I think that will bring the bottom up”. Another administrator expressed significant doubts regarding the true impetus behind the plan, however, sharing that “the reality is a lot of people are doing it because they want the scores to go up for the school. They don't want to teach lower level kids”.

In Beacon Hills and Pine Coast, there were still some participants – particularly teachers – who expressed similar doubts about “mixing” students of differing performance groups. That said, however, there was a greater propensity across both schools for educators to see

opportunity in classes serving students of “mixed” ability levels. Several participants in Beacon Hills, for instance, expressed that, while heterogeneously grouped classes presented significant challenges, the faculty were willing to come to grips with those challenges in the name of promoting student learning. Several educators in Beacon Hills referenced engaging in mixed-ability grouping as a strategy for motivating greater achievement among lower-performing students. One teacher in the school gave an example, reporting that

the way I did the seating chart, and I think other teachers do as well, I figure out who are my high score FCAT students. I put them all in one row, and you have the students that you know have lower FCAT scores, so when you do partners, you are partnering up somebody that's stronger with somebody that's weaker.

Participants in Pine Coast indicated that there were systemic efforts to promote mobility of students upward through the hierarchy of courses in the school. In particular, the school supported a limited cohort model, accompanied by a course designed to support students acquisition of academic skills, intended to promote higher-level course taking among students. In both schools, administrators reported that they placed an emphasis on scheduling outside of the district matrix, generally in an effort to “push” students in an upward trajectory. An administrator at Beacon Hills, for example, shared that “we, at Beacon Hills, didn't like their charts. We felt they were not pushing the kids. They wanted kids at other schools to have same opportunities that they have here, but the charts were gearing our students a little low”.

Educators' Framing of Expectations

Across all four schools, participants shared that the “knowledge” they derived about students from academic performance data structured their perceptions of students’ potential and expectations for their success. In general, however, educators in higher and lower value-added schools framed their data-based expectations for students in very different ways. In our lower value-added schools, both administrators and teachers expressed frustration with their “lower

performing” students, often ascribing students placed within these categories with traits inimical to the promotion of student achievement, like ignorance or apathy. In our higher value-added schools, some teachers expressed that working with students in the “lower” performance bands could be a challenge, but they generally counter-balanced this perspective with an overriding belief that all students could learn, and that it was the job of the faculty to make it happen.

Several participants in Cypress Cove and Bay Mountain – the schools identified as “lower performing” by our value added analysis – expressed a perception that students grouped into the “lower” academic performance bands, including the “bottom 30%”, possessed more limited academic capacity than their “higher” performing peers. A teacher in Cypress Cove shared, for instance, that

that is my lowest performing class, and you can see by their data most of them are level one and two on FCAT, very low readers, way below grade level. Then, you know, giving a lot of-- you can only give them, expect them to do so much [...] unfortunately those kids, you can expect very little. Not every kid, but as a general, when the class is that low, they won't do a lot based on data.

A teacher in Bay Mountain expressed similar sentiments, lamenting that

As much as I am trying to reach these kids, and get them involved academically, I feel like I am up against-- I'm trying to push back the glacier, it feels like sometimes. And I question whether these kids even belong in a traditional high school setting. Aren't these kids perhaps better served in a vocational setting. Do they need Macbeth? What they really need is some kind of skill. They are not going to go to college. Why are they sitting here every day, when there might be a better way to get to these kids and get them prepared for the real world.

Interestingly, a teacher in Cypress Cove spoke directly about the dissonance between the “picture” that the performance data painted of her students and her own intrinsic desire to believe in their potential. She shared that

I mean, I am from the class of Harry Wong, where you are not supposed to look at this [performance data], and you are supposed to think that everybody can do it. They want you to look at this constantly, and if I look at this constantly I am going to think they are

stupid, where I always think my kids can do it [...] but constantly looking at this makes me think they are retarded.

Another commonly referenced attributed of the “bottom 30%” related to their apathetic approach to academic effort. One teacher, for example, reported that one of the primary stumbling blocks for Cypress Cove lay with the relative lack of effort among such students:

We made as many gains across the board as we could, except for that deep 30 percent that is just engrained in mediocrity, low scores, and not giving a care. That ended up being our problem, was not enough of them had gains. So in the end, the shared focus group is that we want to get better. The problem is that the students aren't in that same mindset. So no matter how hard we teach these kids, we are leading horses to water and they want soda.

Another teacher in Bay Mountain shared this sentiment, indicating that she perceived that students in the “lower” academic tracks were simply less willing to work than their higher-performing peers: “I deal with a lot of kids who are low level, and because they are low level, not always but usually, their attitude will reflect it, their behavior will reflect it. I have students who don't care. Don't try. Who have more fun disrupting the class themselves, instead of [learning]”.

In general, educators in Beacon Hills – one of the schools identified as “higher performing” by the NCSU’s value-added model - expressed qualitatively different perceptions of the potential held by students in the “lower” academic tracks. First, educators in Beacon Hills – both administrators and teachers - emphasized a fundamental belief that all students could learn, even if some students were more challenging to teach than others. Moreover, educators in the school asserted that students –regardless of their perceived ability “level” – should be held to high standards. An administrator in Beacon Hills for example, asserted that

the goal here [...] is for them to be in the most rigorous courses possible. I just read an article that talks about that, where kids the expectation level of themselves and the parents have on them slights the kids. We are the ones that have to push and say, after

looking at your scores that you are capable of being successful in that class. So we have to promote that with them.

Another aspect of educators' framing of expectations for students that differentiated Beacon Hills from our lower value-added schools centered on where teachers and administrators placed the locus of control over students' learning. A teacher in Beacon Hills framed this notion well, asserting that:

I have had the experience of teaching both, but for the last two years it's been exclusively our level one and two. It's funny. The challenges I see are more social than academic, because the academic piece, it's not the root of the issue. It's a symptom of the social issues that are taking place [...] we can work on that piece, the academic piece, as long as you have a strong teacher in front of them, that knows about how to reach them, that knows about what that looks like, and how to measure that, and how to create. As long as you have a strong teacher, the academic piece can be taken care of.

A member of the school's administration agreed, emphasizing that Beacon Hills's leadership expected faculty to scaffold students' educational experiences, but not to temper their expectations for student success: "they may have to adjust their lessons. They may have to adjust the rate in which they deliver lectures. They may have to adjust the amount of assignments, but it shouldn't be to decrease their expectations, or performance of the students. It may just be the rate at which they do it".

Responses from Pine Coast were more mixed than in Beacon Hills across both these dimensions; in many ways, the participants' framing of expectations for student performance seemed somewhere in the middle of the continuum between our case study schools. There were still teachers, for example, who expressed poor expectations for students framed by data or course-level as being "lower-performing". That said, a core group of educators in the school - comprised of several administrators and teachers - expressed perceptions similar to those evoked in Beacon Hills, and this differentiated Pine Coast from our lower value-added schools. A

teacher in Pine Coast, for instance, asserted his belief in the importance of holding high expectations for all students:

I really think the students will rise to the occasion. So I would like to see that happening. I would like to see less of bringing a subject down to meet those needs. I would rather see the students be brought up. That's when I get frustrated, if I feel like we have to dumb down a situation, I am not happy with that at all. That's my success for the school is to keep it at a high playing field.

An administrator in the school shared similar sentiments, reflecting that “when with the kids that come to the table with numerous issues, my goal is for them, whether they graduate or not, is to be a successful citizen and I feel like the reward is when they return after they graduate [...] and tell me how great they are doing”.

Educators' alteration of practice

Across all four case study schools, educators reported that their perceptions regarding students' ability and potential – framed and shaped by student performance data – translated into real changes in instructional practice. Administrators across all four schools reported that they expected teachers to “differentiate” their instructional practice in order to meet the varying capacities and needs of students. An AP in Cypress Cove, for example, shared that “in our school a lot of our classes are mixed, meaning we have level one students with level four students. I want to make sure that teachers are differentiating instruction, making sure they know how to differentiate instruction. We talk about that a lot”. Teachers, by and large, reported that they understood this expectation, and worked to meet it – a teacher in Pine Coast, for example, shared that

it's the lower level kids, the bottom 30 percentile, as they are labeled, according to their test scores. It's always been a challenge because you are always looking to how can I reach out to this kid. Each kid brings a different personality, and different strengths and weaknesses as far as comprehension, just their overall intellect. Each kid brings a different approach.

The ways in which “differentiated instruction” was defined and operationalized, however, varied widely between school contexts.

In Cypress Cove, nearly all teachers reported that they were cognizant of the need to “differentiate” their instruction, but few were able to clearly articulate how. The assistant principal quoted above recognized the disparity between the need for “differentiation” and capacity for doing so in her staff, sharing that “we talk about differentiated instruction, and insuring that teachers are teaching all different levels within that classroom, but we are not sure the teachers knows what differentiated instruction is”. When teachers in Cypress Cove did describe the process by which they “differentiated” their instruction, they shared a few strategies – including reducing the level of rigor to meet “lower-level” students’ capacity, adjusting the pace of the class or repeating material. Almost universally, educators in the school noted that they were expected to know what “level” their students were at, and to tailor their instruction accordingly, but there was little consistency in how practitioners described doing so.

Bay Mountain differed from Cypress Cove in that teachers were much clearer and more consistent in their definition of what “differentiated instruction” meant – namely, that they responded to the needs of “lower performing” students by reducing the instructional rigor of their classes. One teacher, for example, contrasted his “upper” and “lower” level courses by asserting that

I am not taking that class that you saw of 10th grade gifted, and we are not spending two weeks on comparing and contrasting. Are we going to compare and contrast? Certainly. We are going to do that at a much higher level. We are not going to go and drill and kill FCAT packets, and we are not going to do work sheets. I am not. That's going to destroy them.

Another teacher in the school framed a similar conception of “differentiation”. She described how she differentiated assessments for her students, sharing that

how I structure my lessons, what my students need, the types of tests I give them, even the vocabulary quizzes, they are different from my honors student to the regular. This quiz I gave you is so simple, it's just matching and then sentences. How easy can you get [...] My honors kids, they have that list on the board, they write down the words, they define it, and they use a sentence all from their brain [...] but I can't give that same quiz to my 10th grade regular kids now, because they are not at the same -- things like that are driven by data.

This pattern – of reducing rigor for students framed as being “lower” performing or of lower ability – held across the several teachers in the school. An administrator in the school seemed to capture the essence of the phenomenon, reflecting that, for “lower” performing students, “I would say their experience is not as rich, perhaps. I see them as experiencing the barebones. We give them what they have to have to get their high school diploma”.

In Beacon Hills, teachers reported an understanding of “differentiation” that was consistent emphasizing the importance of coming to “know” students’ “strengths and weaknesses” through the use of performance data. An administrator in the school expressed that this nuanced perspective on “differentiation” was a focus of development in the school, noting that staff worked together to

identify some of their level one or level two readers, or level one or level two math students. They identify that and turn that into us, and develop their professional growth plans based upon the data that they receive, and what the kids need, whether it's reference or research or benchmark that the these kids are deficient in. That's who the teacher needs to identify, and needs to have a strategy for.

Several teachers in Beacon Hills corroborated this; one described this process of understanding students’ “strengths” and “weaknesses”, sharing:

In the very beginning of the school year we will have a meeting where we will get computer carts out and we will pull up data from our students, and it's their FCAT scores, to see whether-- the FCAT scores can range from one to five. So we will see are they scoring-- how well they are doing on their FCAT scores, and also see what their other test scores are like. What they like us to do is see what skill they are lacking. So let's say my first period class I look at their data and I can see that for math they are lacking Algebra skills, so they would like us to implement more Algebra skills to help them.

On the whole, Educators in Pine Coast responded similarly to those in Cypress Cove, in that there was not a strong, consistent conception of what it meant to “differentiate” individual instructional practice across participants. Several teachers referred to making a concerted attempt to “get to know” students in the identified as being “lower performing”, and to attempt to construct stronger, more personalized relationships with such students. Others indicated that, like Cypress Cove and Bay Mountain, they responded to signals they received from data about student ability by altering the instructional rigor of their classes. One teacher, for example, shared that

my lowest level kids are my fourth hour class. Every single one of those are in the 30th percentile. So the information, or work, I give them is a little lower. Not lower. I don't like using lower. I like using more adaptive to what their needs are. The key that we are using this year is referencing and research. I will emphasize more records and research on that test, or that quiz, or that writing. Whereas, let's say my fifth hour, they are a little higher, so I won't use as much referencing and research, I will probably use more cognitive, or more theory associated with what they need to do.

Unlike Cypress Cove and Bay Mountain, however, a strong school-level system existed in Pine Coast for meeting the challenges presented by the varying “ability levels” of students – through systemic targeting and assignment of academic supports, like pull-outs and tutoring, for students labeled as part of the “bottom 30%”.

In addition to the interviews referenced above, the National Center on Scaling Up Effective schools conducted numerous and comprehensive classroom observations in each of the case study schools using the CLASS-S observation tool¹. In general, analysis of classroom observations revealed few systematic differences between higher and lower value-added schools; there were, however, differences between the schools in instructional quality between course levels (e.g. honors and regular courses) across schools (Rutledge, Cohen-Vogel & Osborne-Lampkin, 2012). Across CLASS-S elements, which included constructs like classroom

organization, instructional support and student engagement, observed quality tended to be lower for students in “regular” courses – as our participants have indicated, generally those students in groups like the “bottom 30%” - in all four schools. These gaps were narrower, however, in Beacon Hills than in any other case study school (Rutledge, Cohen-Vogel & Osborne-Lampkin, 2012).

Educators’ intended outcomes

In all four case study schools, participants – both administrators and teachers – rationalized the practice of “labeling” or “grouping” students, and their efforts to differentiate students’ experiences, in a variety of ways. In Cypress Cove and Bay Mountain, the two schools identified as being lower performing in the NCSU’s value-added analysis, educators tended to frame their efforts as a means for overcoming the challenge of maximizing educational productivity – generally in response to accountability demands. Several framed their work in terms of “moving” students across the performance spectrum. An AP in Cypress Cove shared, for example, “that reading is an area we are hoping to improve on primarily with our lowest quartile students. The state's expectation is that at least 50 percent of the lowest quartile students will make learning gains, and we didn't make that mark last year”. A teacher in the school lamented this emphasis on grouping, productivity, and “movement”, asserting that

[The] 30 percentile is what we aim for, because that's where they need to make gains to keep the school at the level you are. You can never achieve the best in the state of Florida because there is-- every time you move up, it's going to move higher. It never stops. There is no top. You may move seven percent this year, but next year you have to move another seven percent to maintain that grade, otherwise you fall down a grade. That's frustrating. They keep asking for more, more, more, and you can only get so much out.

Educators in Bay Mountain used similar language but were, in many ways, more evocative regarding the potential consequences of the school’s cultural focus on productivity for

equity of achievement among students. A teacher in Bay Mountain, for example, reflected on the principal's focus for the school: "he wants us to do well. And, I feel like he-- last year, not so much this year, last year there was sort of a desperation to do everything possible to remedy our D, because we were a D school. So, that desperation sometimes isn't always productive. Doesn't always lead to productivity". Other educators in the school were more blunt; one teacher asserted that: "the principal seems to be concerned about results on the standardized tests, which is what the district put on his shoulders, and every effort in the school is toward the test. It seems that way, and we all take it as that way. It's just FCAT". A final teacher shared that "all he talks about is numbers. The kids to him mean numbers. I guess for him all he wants is to keep his job and make sure [...] we meet AYP, so he looks good. I think that's all he wants, honestly".

Several teachers expressed concern that this rigid focus on the part of school leadership was a detriment to students outside of ability groups targeted by accountability policies. One teacher shared the perspective of several others, relating that

We are focusing so heavily on the kids that are in the lowest quartile, that we abandon the kids in the middle and the top. We just expect them to somehow figure out how to go to college by themselves. Not in my class particularly, but I think that's in general a kind of feeling that's on campus. We don't have a huge movement towards higher education, and the value of it, and the positive aspects of it.

Educators in Beacon Hills and Pine Coast, the two higher performing case study schools, shared similar concerns regarding the pressures of accountability. Several educators across the two schools, however, expressed a shared value in ensuring that all students – regardless of ability “group” or level – learned and were successful. An administrator in Beacon Hills captured the essence of this shared value, stating that “I don't differentiate my students. I don't say, you are an AP kid, or you are-- all students-- it's our job to education all students. No matter what”. A teacher in Beacon Hills expressed similar sentiments, asserting that “I am teaching them. I am teaching them what they should be taught any way. If I see a kid struggling, I am going to assist

them without having to know they got a one on their reading”. Another shared that this strong focus on equal opportunity for achievement was a strong, leadership driven, focus in the school, sharing that she understood the principal’s goal to be that “he wants everyone to do their job. He wants to make sure that all of the kids get the advantage to graduate”.

Several teachers in Pine Coast expressed similar perspectives. One, for example, asserted that “I am not concerned about what letter grade we get. I am more concerned about whether we are meeting the needs of our students. I think if we are able to do that, and they are able to achieve, and able to go to college and pass that doggone FCAT test, then I think we have done a good job”. Another expressed his personal value on equity, sharing that “I don't like when someone says to me, someone has to do the garbage. I don't like when I hear a teacher make a comment like that. They get it, they get it. They don't, they don't. It's very discouraging”. Another teacher shared her perspective, asserting that the teachers’ focus in the school should not be bound by certain conceptions of which students should be “moved” – “getting each and every kid-- that's our job, to bring the best out of them”.

Discussion

These findings present a fascinating picture of how four case-study schools appear to be negotiating the challenges presented by increasingly varied and complex student populations, mounting pressure to respond to comprehensive student performance data and school cultures still grappling with entrenched beliefs regarding the nature of student ability and potential. In our lower value-added case study schools – Cypress Cove and Bay Mountain – educators expressed a keen awareness of the accountability frameworks surrounding them. Their response to these pressures emphasized the need to compartmentalize – and, in many cases, exercise a kind of triage around – those students that they felt an urgent need to “move” in order for their schools to

make the grade. The growing wellspring of performance data surrounding each school provided a seemingly objective framework by which they could do so – one that, in many cases, seemed to reify their perceptions regarding the ability and potential of their students. With these data, educators in Cypress Cove and Bay Mountain were able to easily categorize their students into groups like the “bottom 30%” or “level ones” and “level twos”. Teachers and administrators in these schools seemed to frame students in these groups as challenges to be overcome; in many cases, apathetic or ignorant weights upon the faculty, preventing them from driving the continual growth and productivity they were mandated to ensure.

In our higher value-added schools – Beacon Hills and Pine Coast – many participants responded differently. In Beacon Hills, especially, educators seemed to embrace a culture based upon a fundamental notion that all students could and should learn. That was not to say that all students should learn in the *same way* or at the *same pace*; the emphasis in Beacon Hills’s school culture seemed to be placed on understanding where students “were” so that teachers and administrators could push them toward the same end-state – graduation and academic success. Within this culture, performance data served as a tool for coming to “know” the “strengths and weaknesses” of students so that educators could meet their needs as individuals, rather than as parts of an ability group. Pine Coast, interestingly, seemed to be a school in a transitional state. While many faculty members seemed to frame their practice in the same way that teachers in Cypress Cove and Bay Mountain did, there seemed to be a core movement in the school toward a culture similar to Beacon Hills’s. This movement focused on using data to “target” support for students, and to break students in the “bottom 30%” loose from the constraints of low rigor and poor expectation, at a systemic level. The effort to shift that focus from the “core” to the level of broad school culture was, however, still underway in Pine Coast.

Looking across our four case-study schools, the experiences of participants in Cypress Cove, Bay Mountain, Beacon Hills and Pine Coast support the case made by Oakes, Wells, Jones & Datnow (1997) and many others – that the culture of a school matters significantly in whether or not systemic efforts to ensure equitable access to educational opportunity, and to break apart inequitable structures that group and “track” students, succeed. These data indicate that the presence of an affirmative and positive culture emphasizing the importance of equity also has significant implications for how our case study schools internalized and made sense of accountability frameworks and performance data. Without the presence of a school culture focused on equity and universal learning – like Beacon Hills’s and (although in many ways still nascent) Pine Coast’s – these forces seemed to open the door for educators to move back toward practices that have, in the past, been shown to negatively impact underserved groups of students.

References

- Anderson, S., Leithwood, K., & Strauss, T. (2010). Leading data use in schools: Organizational conditions and practices at the school and district levels. *Leadership and Policy in Schools, 9*(3), 292-327.
- Argys, L., Rees, D., & Brewer, D. (1996). Detracking America's schools: Equity at zero cost?. *Journal of Policy Analysis and Management, 15*(4), 623-645.
- Brophy, J., & Good, T. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology, 61*(5), 365-374.
- Burris, C., Heubert, L., & Levin, H. (2006). Accelerating mathematics achievement using heterogeneous grouping. *American Educational Research Journal, 43*(1), 105-136.
- Clotfelter, C., Ladd, H., & Vigdor, J. (2005). Who teaches whom? race and the distribution of novice teachers. *Economics of Education Review, 24*, 377-392.
- Cohen-Vogel, L. & Harrison, C. (in press). Leading with data: Evidence from the National Center on Scaling Up Effective Schools. *Leadership and Policy in Schools*.
- Cohen-Vogel, L. (2011). "Staffing to the test": Are today's school personnel practices evidence based? *Educational Evaluation and Policy Analysis, 33*(4), 483-505.
- Congar, D. (2005). Within-school segregation in an urban school district. *Educational Evaluation and Policy Analysis, 27*(3), 225-244.
- Corbin, J. M., & Strauss, A. L. (2008). *Basics of qualitative research, techniques and procedures for developing grounded theory*. Sage Publications, Inc.

- Firestone, W. & González, R. (2007). Culture and processes affecting data use in school districts. *Yearbook of the National Society for the Study of Education*, 106: 132–154.
- Gamoran, A. (1992). The variable effects of high school tracking. *American Sociological Review*, 57, 812-828.
- Gamoran, A., & Mare, R. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement or neutrality. *American Journal of Sociology*, 94(5), 1146-1183.
- Gamoran, A., Nystrand, M., Berends, M., & LePore, P. (1995). An organizational analysis of the effects of ability grouping. *American Education Research Journal*, 32(4), 687-715.
- Goldring, E., Porter, A., Murphy, J., Elliot, S., & Cravens, X. (2009). Assessing learning-centered leadership: Connections to research, professional standards, and current practices. *Leadership and Policy in Schools*, 8(1), 1-36.
- Good, T. (1987). Two decades of research on teacher expectations: Findings and future directions. *Journal of Teacher Education*, 38, 32-47.
- Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership*, 60(5), 6-11.
- Hallinger, P., & Murphy, J. (1986). The social context of effective schools. *American Journal of Education*, 94(3), 328-355.
- Halverson, R., Grigg, J., Prichett, R., & Thomas, C. (2007). The new instructional leadership: Creating data-driven instructional systems in school. *Journal of School Leadership*, 17(2), 159-194.

- Hoffer, T. (1992). Middle school ability grouping and student achievement in science and mathematics . *Educational Evaluation and Policy Analysis*, 14(3), 205-227.
- Ingram, D., Louis, K., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record*, 106(6), 1258-1287.
- Klem, A., & Connell, J. (2004). Relationships matter: Linking teacher support to student engagement and achievement. *Journal of School Health*, 74(7), 262-273.
- Lee, V., & Bryk, A. (1988). Curriculum tracking as mediating the social distribution of high school achievement. *Sociology of Education*, 61(2), 78-94.
- Louis, K., Leithwood, K., Wahlstrom, K., and Anderson, S. (2010). Learning from leadership: investigating the links to improved student learning. Final report of research findings. Report Commissioned by The Wallace Foundation.
- Lucas, S., & Berends, M. (2002). Sociodemographic diversity, correlated achievement, and de facto tracking. *Sociology of Education*, 75(4), 328-348.
- Natriello, G., Pallas, A., & Alexander, K. (1989). On the right track? curriculum and academic achievement. *Sociology of Education*, 62, 109-118.
- Oakes, J. (1992). Can tracking research inform practice? technical, normative, and political considerations. *Educational Researcher*, 21(4), 12-21.
- Oakes, J., & Guiton, G. (1995). Matchmaking: The dynamics of high school tracking decisions. *American Educational Research Journal*, 32(1), 3-33.

Oakes, J., & Lipton, M. (1992). Detracking schools: Early lessons from the field. *The Phi Delta Kappan*, 73(6), 448-454.

Oakes, J., Ormseth, T., Bell, R., & Camp, P. National Science Foundation, (1990). *Multiplying inequalities: The effects of race, social class, and tracking on opportunities to learn mathematics and science*. Santa Monica, CA: The RAND Corporation.

Oakes, J., Wells, A., Jones, M., & Datnow, A. (1997). Detracking: The social construction of ability, cultural politics, and resistance to reform. *Teachers College Record*, 98(3), 483-510.

Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Sage Publications, Inc.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. New York: Hold, Rinehart & Winston.

Rutledge, S. A., Cohen-Vogel, L., & Osborne-Lampkin, L. (2012). *Identifying the Characteristics of Effective High Schools: Report from Year One of the National Center for Scaling Up Effective Schools* (Research Report) (pp. 1–109). Retrieved from http://www.scalingupcenter.org/data/files/gallery/ContentGallery/NCSU_Identifying_the_Characteristics_of_Effective_High_Schools_Sept_2012_Final.pdf

Slavin, R. (1987). Ability grouping and student achievement in elementary schools: A best-evidence synthesis. *Review of Educational Research*, 57(3), 293-336.

Wells, A., & Oakes, J. (1996). Potential pitfalls of systemic reform: Early lessons from research on detracking. *Sociology of Education*, 69, 135-143.

Yonezawa, S., Wells, A., & Serna, I. (2002). Choosing tracks: "Freedom of choice" in detracking schools. *American Educational Research Journal*, 39(1), 37-67.

End Notes

¹For more information regarding the CLASS-S and the observations, see Rutledge, Cohen-Vogel & Osborne-Lampkin, 2012.