



## RESPONSE TO PROGRAM SCORECARDS

Wilmington University has been a strong supporter of Delaware’s school reform efforts for many years, and continues to be a strong supporter of comprehensive and meaningful assessment strategies that contribute to improving the quality of teaching and learning at all levels – pre-school through college. Because of our long-time commitment to outcomes assessment, we have been fully supportive of Delaware’s initiatives related to assessing and improving Educator Preparation Programs in this state. We continue to hold that view.

It should be noted here that Wilmington University’s most recent accreditation review (in 2015) by the Middle States Association of Colleges and Schools resulted in full accreditation with NO recommendations for improvement. In addition, the College of Education is fully accredited by CAEP (formerly NCATE) and our individual degree programs are Nationally Recognized by their Specialized Professional Associations.\* In fact, the most recent NCATE Accreditation Review (in 2013) resulted in full accreditation through 2020, and a rare commendation for our sophisticated and effective outcomes assessment system. Because of this, we have been asked to share our assessment philosophy and architecture at several national education conferences, with other IHEs in this state and others, and with the CCSO (Council of Chief State School Officers).

We are firm believers that assessment strategies must be focused on *outcomes* more so than on *inputs*. They must also measure what is important, even though that can be very difficult, and do it in such a way that the results can be used to make meaningful data-driven program changes and improvements. The old adage among assessment and program evaluation specialists is: “We often measure what is unimportant, and fail to measure what is important.” In the case of the Delaware Department of Education’s current effort to measure and report on the quality of Teacher Preparation Programs via Educator Preparation Program Scorecards, some of the items measured can be categorized as *unimportant* and others could be considered irrelevant and invalid.

We are particularly concerned with four domains reported on the Program Scorecards: Recruitment, Candidate Performance, Placement, and Retention.

### RECRUITMENT

The first domain measured and reported on is “Recruitment.” The domain is described as a program’s “ability to cultivate a diverse, accomplished student body with the potential to become outstanding educators.” A major component of that measurement is “Average SAT Score of Incoming Class.” There are several problems with this metric.

1. The State of Delaware does NOT require that prospective public school teachers take the SAT or, if they have taken it, does not require applicants to provide SAT test scores as a condition of initial licensure or certification or employment as an educator. Consequently, the use of SAT scores to evaluate the quality of a preparation program makes little sense, since the state does not use such scores to evaluate a candidate’s potential or to evaluate the effectiveness or quality of its educators.

---

\*The scorecard fails to include this information, even though the final page of the report has a section titled ACCREDITATION. The notation for Wilmington University CAEP and SPA accreditation is **NA**.

2. Another problem with using only SAT scores to measure this domain is that Delaware legislation (SB 51, signed into law in June 2013) specifically lists three acceptable measures of “general knowledge” as acceptable options for admission to teacher preparation programs. The exact language is, “Demonstrate mastery of general knowledge, including the ability to read, write, and compute, by achieving a minimum score on a standardized test normed to the general college-bound population, such as Praxis, SAT, or ACT, as approved by the Department.” However, the only standardized test for which the Department of Education has set a minimum score is Praxis I (now Praxis Core Series for Educators). While Delaware does accept SAT scores (Verbal and Mathematics) in lieu of Reading and Mathematics Praxis scores, the state has NOT established an overall cumulative SAT score requirement. Consequently, there is no SAT standard by which an individual student or an educator preparation program can or should be judged.
3. The majority of high-school students in this country do not take the SAT, because it is not required in most states. In fact, more than 800 colleges either do not require SAT test scores for admission, or are “SAT Optional” schools, including some that are very selective (e.g., Bennington, Brandeis, George Washington University), and some that are outstanding educator preparation institutions (e.g., Western Governors, Arizona State). For a complete listing of those schools, go to <http://www.fairtest.org/university/optional>. This means that many applicants for admission to college have never taken the SAT, and have no scores to report. A major flaw in the Recruitment scorecard involving SAT scores is that the number of scores being reported is not stated. In other words, an incoming class of candidates might have as few as 10 students whose scores are being reported, but the actual class could contain many times that number of students. Consequently, an entire program could be judged based on the average SAT score of a very small, percentage of its students.
4. Although Delaware now provides the opportunity for all of its students to take the SAT, this is not the case in most states. In fact, the ten states with the highest average SAT scores actually test 6% or fewer of their students. (North Dakota, ranked #2 in the country, tests only 2%). Again, this means that a sizable number of students applying for admission to colleges across the country have no SAT scores at all. The only metric that would be even marginally acceptable in terms of measuring recruitment would be one that could be applied equitably to ALL students entering a program, not just a handful.
5. The SAT score average reported on the Program Scorecard is relatively meaningless simply because it is an average of a set of scores. A metric such as that is useless in evaluating a program’s quality or even an individual candidate’s possible weaknesses and strengths. For example, an SAT test-taker could obtain a very high score on the mathematics section, but a relatively low score on the verbal section, and receive an overall “average” SAT score that looks pretty good. However, there is some research suggesting that highly “literate” teachers are more effective as teachers, especially at the elementary school level, but only the verbal portion of the SAT is correlated with overall literacy.
6. We know that many first-generation, poor, and minority students do less well on the SAT than test takers from some other demographic groups. There are several reasons for this, not the least of which is that students from low-income neighborhoods tend to go to schools that are of lower quality than those in more affluent neighborhoods, and schools in low-income neighborhoods usually offer far less in terms of challenging curriculum options along with the extra-curricular and leadership opportunities that are known to contribute to post-secondary school success. Therefore, using the SAT as a measure of “recruitment” effectiveness is an obvious contradiction, because it mitigates against “cultivating a diverse, accomplished student body.”
7. An extensive, three-year study by Hiss and Franks (Defining Promise: Optional Standardized Testing Policies in American College and University Admissions, 2014) found that whether or not students submitted SAT or ACT scores as part of the admissions process had no significant impact on their GPAs and graduation rates.

8. According to a report issued by the National Center for Teacher Quality (NCTQ, 2004. *Increasing the Odds: How Better Policies Can Yield Better Teachers*), an organization whose work is often quoted by officials in the Delaware Department of Education, “measurable teacher attributes like SAT scores and experience account for only a small portion of why some teachers are more effective than others.” That same report states that “most of what makes a teacher effective are the *soft* personal attributes that are much harder to measure.” The report cites work done by the Haberman Foundation and Teach for America showing that personal attributes are extremely important components of an educator’s success. Those researchers found that effective educators were:
  - a. Persistent: could survive within a bureaucracy.
  - b. High-Achieving: had a history of success no matter what the endeavor.
  - c. Responsible: instead of blaming others or circumstances, took full responsibility for achieving a positive outcome.
  - d. Critical thinkers: reflected on the linkages between cause and effect instead of simply reacting to the effect.
  - e. Organized: were able to juggle multiple projects and tasks successfully.
  - f. Motivating: were able to influence and motivate others to action, as evidenced by effective leadership in extracurricular activities such as student-run organizations or athletic teams.
  - g. Respectful: assumed the best about people, especially people in low-income communities.
  - h. Advocates: they shared the mission and goals of the organization, e.g., they believed in trying to eliminate educational inequities.
9. Consequently, Wilmington University believes that dispositional characteristics such as those described in #7 above are more important than test scores when making admissions (and program completion) decisions. The importance of those characteristics was also affirmed by CAEP’s recent decision (in June 2013) to adopt Standard 3.3, which directs preparation programs to establish and monitor attributes and dispositions that candidates must demonstrate at admission and during their programs.
10. Finally, the mission of Wilmington University is clear. We are “committed to excellence in teaching, relevancy of the curriculum, and individual attention to students. As an institution with admissions policies that provide access for all, it offers opportunity for higher education to students of varying ages, interests, and aspirations.”
11. For all of the reasons cited above, we cannot support the use of SAT scores as a condition of admission or as a metric to rate the quality of our students or our educator preparation programs.

### **CANDIDATE PERFORMANCE**

This part of the Scorecard uses *Praxis* scores (from the ETS Praxis Core Series for Educators or Praxis I) to evaluate the CANDIDATE PERFORMANCE domain, along with a Performance Assessment. The scorecard’s definition of candidate performance is “the educator preparation program’s ability to prepare aspiring educators with the knowledge and skill to be first-day ready ...”

The Wilmington University College of Education has adopted the PPAT (Praxis Performance Assessment for Teachers), a rigorous and complex assessment tool developed by ETS, which will be required for all teacher preparation candidates completing their programs after July 1, 2016. Our programs have been restructured in such a way as to fully prepare our candidates for success on this assessment, and we support its use. However, Delaware has not yet set a minimum passing score requirement for the PPAT, even though the Professional Standards Board has had more than a year to do so.

A major problem with this domain is the use of Praxis I scores as a measurement of a program’s ability to “prepare aspiring educators with the knowledge and skill to be first-day ready ...” The scorecard uses average Praxis I scores of graduates who have worked in a Delaware public school to assess this domain. This approach has the same flaw

as the Recruitment metric, wherein the scorecard reports an average of an average set of scores, which reveals little about candidates' general knowledge in each of the tested areas.

An even more significant flaw with this metric is that the Praxis I exam is a test of basic general knowledge in reading, mathematics and writing and has nothing to do with college-level content knowledge or professional skill. The General Assembly's intent in SB 51 was that Praxis I could be used in making admissions decisions and clearly defines Praxis as an example of an acceptable program entry requirement (see #2 under the Recruitment comments). Nowhere does the law imply or require that Praxis I scores be used as a measure of candidate performance within a program or following graduation. In fact, the law describes Praxis scores as acceptable measures of the college-bound population's "ability to read, write and compute," rather than candidates' performance after they have arrived at college.

For that reason, we require that all of our candidates achieve scores on all Praxis I subtests that meet or exceed Delaware's licensure/certification requirements as a condition for ADMISSION to any of our educator preparation programs (even though Delaware accepts composite Praxis test scores and even permits aspiring educators to FAIL one part of the exam).

We suggest that if the Department of Education wants to use a standardized test reflecting the "preparation program's ability to prepare aspiring educators with the knowledge and skill required to be first-day ready," the ETS Praxis Subject Assessments (formerly ETS Praxis II exams) should be used instead. According to ETS, "these tests measure subject-specific content knowledge as well as specific teaching skills needed for beginning teaching," precisely what the Scorecard is attempting to measure. For that reason, Wilmington University has always required its candidates for educator preparation degrees to complete and pass (using established Delaware minimum test score standards) the Praxis II assessment that corresponds to a candidate's preparation area (elementary education, special education, counseling, reading, etc.) ... and to do so BEFORE being permitted to enroll in the final clinical experience (internships, student teaching, etc.). For information about these exams, see: [Praxis Subject Assessments information](#)

## **PLACEMENT**

The Placement domain includes two metrics. The first metric describes the rate at which program graduates begin working in public education in Delaware. The scorecard shows that Wilmington University is at or near the top of the rating scale in comparison with other Delaware IHEs, something we have known for many years, and something about which we are quite proud.

However, the second metric is "placement rate in Delaware high-needs schools." While our graduates again score near the top of the range in this category, the metric itself is completely invalid as a way to measure a program's effectiveness, because educator preparation programs have absolutely no control over where beginning teachers are placed. Those decisions are made entirely by the employing school districts, and are influenced by collective bargaining agreements and seniority. However, since it is well-known that school districts tend to assign more experienced teachers to more "desirable" classroom settings, we try to prepare our graduates for the most challenging teaching situations possible, and we purposely provide our candidates with field experiences in a variety of urban, suburban, and exurban settings. As can be seen from our "Retention" ratings below, that approach is proving to be very effective.

It should also be noted that our Birth-Grade 2 Program was given a very low placement rating in the current scorecard. However that rating is biased and invalid, because almost none of our graduates from that program seeks employment in public school settings at the pre-school (age 4) level. Nearly all of our candidates in the 0-2 program are preparing for (and seek) work with very young children (infants and toddlers), and to my knowledge there are no such programs offered in Delaware public schools. Consequently, for DOE to evaluate a program using a placement metric involving the percentage of our graduates working in public school classrooms serving pre-schoolers (age 4-5) is completely misguided and unfair, and in no way reflects the quality of the program.

## **RETENTION**

The retention metric attempts to measure an educator preparation program's record of preparing graduates who continue to serve in public education in Delaware, and does so by reporting on the percentage of graduates who remain in their positions beyond year one, as well as those who continue working in Delaware public education beyond year three.

Again, Wilmington University graduates score at or near the top of the range when compared with graduates of other IHEs in the state. In fact, the ratings indicate that 91-96% of our teacher preparation graduates continue as Delaware educators beyond year three, indicating that they are achieving high levels of success. However, the problem with this rating is precisely the same as described in the "Placement" domain. An educator preparation program has no control over an educator's decision to continue working in a Delaware school. Those decisions are made by the individual educator, for a broad variety of reasons, not the least of which is the wide pay differential between Delaware and surrounding states (with Delaware being quite low), and working conditions in Delaware schools. In fact, our experience tells us that it is the latter (poor working conditions, lack of professional autonomy, a burdensome bureaucracy, etc.) that contribute most to educators electing to leave their positions. Consequently, to judge the quality of a program based on factors over which the program has no control is completely unfair.

Another important factor over which IHEs have no control has an impact on both Recruitment and Retention, and that is the starting salary a new teacher can expect to earn (assuming that the new graduate is able to find a teaching position in a highly competitive job market). In this category Delaware compares poorly with the other states in the region. For example, Delaware's starting salary for teachers is \$3,000 to \$9,000 less than those in NJ, PA, and MD. Moreover, the average salary of all teachers in those three states is \$5,000 to \$9,000 higher than those in Delaware. Projected over the first 10 years of teaching, the year-to-year differentials remain about the same in each state, meaning, for example, that a beginning teacher in NJ is likely to make nearly \$100,000 more in salary over the first 10 years of teaching than a teacher in Delaware (Source: TeacherPortal.com; NEA). Such differences represent a powerful influence in determining where a candidate decides to enter a preparation program, where the graduate chooses to work, and the trajectory of an educator's career.

## **DATA QUESTIONS**

A general concern we have is the lack of information provided in the Program Scorecards related to the numbers of educators being reported on in each category. In some cases the data make little sense. For example, the Scorecard for Primary Education has an overall rating of Tier 1, the highest rating, yet there are no graduates reported since 2012, and this program was phased out when K-4 certification was eliminated. Another program, the M.Ed. in School Leadership, a degree which contributes to DE certification, and for which we submitted several hundred data sets, is not rated at all. Other data are missing from the Scorecard final pages even though those data were also provided to DOE.

## **SUMMARY**

We reiterate our position that a determination of program quality should be based on outcomes, not inputs. For example, we don't object to the factors used to rate the "Graduate Performance" domain, because those scores are based on actual performance ratings by supervisors, on direct observations of performance, and on student growth and improvement measures (derived from achievement test scores). We feel that these are appropriate, evidence-based outcome measures, and we support them.

We support the state's efforts to evaluate and report on the quality of educator preparation programs in the state. However, we want the metrics used in determining quality to be valid, reliable, and free from bias. The state's first "scorecard" suffers from weaknesses in each of those areas.

We strongly suggest that this first program evaluation effort be considered a field-test, and that we now work collaboratively (with assessment specialists and representatives from all of the state's IHEs and the DOE) to revise and improve the Scorecard's metrics in such a way as to ensure validity and equity. This is something that was NOT done in the preparation of the initial scorecards. In fact, there were only a few information sessions held with IHEs prior to the release of the current reports (and no work sessions), and the current report does not reflect the suggestions and concerns that IHEs expressed at those sessions.

Finally, some of the problem areas described in this document might have been avoided had the Department of Education prepared (in collaboration with IHEs) and published the "Guide," as promised more than a year ago, and which is required by the following section of Regulation 290:

**6.1.4 Units shall compile and report data for each graduating cohort and for the most recent five (5) years of Program cohorts on a selection of metrics. Metrics, as outlined in the Guide ...**

As of this date the Guide still has not been received.

John Gray, Dean  
College of Education  
Wilmington University  
August 31, 2015