Differentiating Military-Connected and Non-Military-Connected Students: Predictors of Graduation and SAT Score

Baker, R.S.[1], Berning, A.[2], Gowda, S.M.[3]

[1] Baker EDM Lab, Inc.

[2] Renaissance Institute

[3] Alpha Data Labs, Inc.

Executive Summary

At-risk prediction and early warning initiatives have become a core part of contemporary practice in American high schools, with the goal of identifying students at-risk of poorer outcomes, determining which factors are associated with these risks, and developing interventions to support at-risk students' individual needs. However, efforts along these lines have typically ignored whether a student is military-connected or not. Given the many differences between military-connected students and other students, we investigate whether models developed for non-military-connected students still function effectively for military-connected students, studying the specific cases of graduation prediction and SAT score prediction. We then identify which variables are highly different in their connections to student outcomes, between populations.

Core recommendations:

1. At-risk prediction models should explicitly consider whether a student is military-connected. Ideally, a separate prediction model and set of indicators should be developed and used for military connected students.
2. Different factors are indicative of at-risk status for military-connected and non-military connected students. These differences should be taken into account when designing data systems to feed into at-risk prediction models. Some core differences are noted in the paper below.

Introduction

Teachers and school administrators have striven to reduce dropout and improve graduation rates for quite some time, but dropout continues to persist in schools until the present day (Wiltz & Seale, 2016). In recent years, efforts to intervene and prevent dropout have begun to leverage advances in predictive analytics, attempting to use mathematical models derived from data to identify students with a particularly high probability of risk. This development has come as the use of predictive analytics in education, referred to at times by the more general terms of learning analytics or educational data mining (Baker & Siemens, 2014), has emerged as an approach to addressing a range of problems in education. Within high schools, increasing numbers of districts are now deploying predictive analytics models, either developed in-house, provided as custom solutions by small vendors such as the Renaissance Institute, or provided by national-level vendors, such as the BrightBytes Early Warning Module.

While very early work attempted to predict school violence (Tobin & Sugai, 1999), attempts to predict graduation only began to emerge as an active area around a decade ago (e.g. Balfanz et al., 2007; Allensworth & Easton, 2007). However, in the last few years, a large number of projects have emerged which attempt to use statistical models and predictive analytics to determine in advance which students will fail to graduate and use these predictions within early warning indicators (EWIs) that help districts allocate resources towards those students most at risk of not completing high school (Allensworth, 2013; Allensworth, Gwynne, Moore, & de la Torre, 2014; Bowers et al., 2013; Carl, Richardson, Cheng, Kim, & Meyer, 2013; Baltimore Education Research Consortium, 2011; Kemple, Segeritz, & Stephenson, 2013; Kieffer & Marinell, 2012). Increasingly, efforts have been made for these predictions to become usable, not just by data scientists or district-level administrators, but by a wider range of school stakeholders, teachers, parents, and students (Bowers et al., 2016). Several districts and research groups have invested heavily in EWI initiatives, with established efforts in cities such as Chicago, Baltimore, and New York City, as well as in more rural areas such as the state of West Virginia, which has a statewide EWI initiative.

However, these models have typically been developed and checked using data from entire district populations of students (or even entire regions of students). While this practice leads to models that are effective for most students, this type of practice carries the risk that models may be less effective for specific sub-groups of students.

In particular, military-connected students may be less well-served than other students by these initiatives. Military-connected students differ from other American students in a variety of important ways; they frequently experience stressors that other children do not, such as the deployment of a parent (De Pedro et al., 2011). These stressors are associated with increased behavior problems during parent deployment and a perception that school staff do not understand the challenges they experience (Mmari et al., 2009). These students also experience a feeling of a lack of connectedness to their school environments, sometimes leading to disengagement from school activities (Chandra et al., 2010). Another major difference between military-connected students and other students is that military-connected students are likely to move house and change school frequently and for different reasons than other children. Although many researchers have argued that changing schools frequently does not directly lead to negative consequences for military-connected students (see review in Palmer, 2008), it does lead to a different experience of schooling and different meanings for indicators within the data that schools have on military-connected students. These differences raise into question whether the same indicators will have the same predictive validity for military-connected students.

In this project, we investigate this question, developing predictive analytics models predicting student graduation and SAT scores for both military-connected and non-military-connected students, in the context of a single Texas district with a significant proportion of military-connected students. We then test whether models developed using data from one population still function equally well on data from the other population. If there is substantial degradation in model quality when a non-military-connected student model is applied to military-connected students, then we can conclude that the current practice of ignoring military-connected status when developing K-12 predictive analytics is problematic, and should be adjusted. We follow this with a discussion of the cases in which specific indicators of potential student outcomes differ for military-connected and non-military-connected students.

Methods:

Data

The data set we used to predict student dropout and violence was generated using data from a highly-diverse small public school district in Texas that serves a population near a major U.S. army base. The high school in this district is just under half White, and just under ¼ Black and just under ¼ Hispanic. Just under half of students are classified as low-income.

Data from 2015-2017 was obtained for this district, so that we could analyze the relationship between student behavior and grades in 2015-2016 (11^{th} grade), graduation at the end of 2016-2017 (12^{th} grade), and SAT scores (which can be taken in either grade). Data from a total of 1,330 students was collected and used in analysis (students leaving the district for reasons other than graduating or failing to graduate – such as a family move due to a change of military assignment or job -- were not included in analysis). Data was provided to the research team in fully deidentified fashion by district personnel; only data on graduation, SAT, military-connected status, disciplinary incidents, grades, course-taking, and attendance was obtained.

309 (23.2%) of the students in the sample were classified by the district as military-connected. 77.4% of the non-military-connected students graduated; 75.1% of the military-connected students graduated. There was not a statistically significant difference in the proportion of graduates between military-connected and non-military-connected students, χ^2(df =1, N=1330)=0.765, p=0.382.

Features Used in Prediction

A total of 215 features (potential predictors for use in data mining) were distilled from the students' 11^{th}-grade data. The following categories of features were generated:

- Features based on the student course grade information (11 features), including features such as average mid-term grade, lowest semester grade in any class, and highest final grade in any class.
- Features based on student attendance (109 features), including features such as how often a student was present or absent from class for specific reasons, including excused absences, unexcused absences, and in-school suspensions.
- Features based on student course-taking (10 features), including features such as how many advanced courses a student has taken, and how many vocational courses the student has taken.
- Features based on the student's disciplinary record (75 features), including features such as the total number of disruptive behaviors recorded for a student, and the total number of dress code violations. Disciplinary records involving school violence were omitted as potential predictors, as we were predicting school violence in the following years.
- Features based on a combination of student course-taking and course grade information (10 features), including features such as the student's average grade for English as a Second Language (ESL) courses, and the student's highest semester grade in any Advanced Placement (AP) course.

Data Mining Approach

Four models were created:

- A model predicting graduation, developed using data from military-connected students
- A model predicting graduation, developed using data from non-military-connected students
- A model predicting SAT, developed using data from military-connected students
- A model predicting SAT, developed using data from non-military-connected students

Separate models were created for military-connected and non-military connected students in order to determine how well models transfer across these populations; in other words, to determine if military-connected students differ from non-military connected students in ways that cause models not to generalize. Contemporary K-12 predictive analytics typically does not take military-connected status into account, and models are developed using data from predominantly non-military connected students, and then those models are used to make predictions about military-connected students as well. Our current analysis will allow us to determine if this practice is benign, or if it is leading to lower-quality analytics for military-connected students.

The question of whether a student graduated or not was treated as a binary classification problem – i.e. we attempted to predict whether or not the student would graduate (1) or would not graduate (0). We attempted to predict graduation using logistic regression, a standard algorithm for predicting binary data which has been frequently found to be effective for predicting longitudinal student outcomes.

The core metric used for assessing the quality of the model predicting graduation was the Area Under the ROC Curve (AUC ROC, or AUC for short) (Bowers et al., 2012). AUC, also referred to in many cases as A', is equivalent to W, the Wilcoxon statistic (Hanley & McNeil, 1982). The Wilcoxon/ A' interpretation of this statistic indicates that it represents the proportion of the time where, if you randomly select one student who will eventually drop out, and randomly select one student who will not drop out, the model can accurately identify which is which. As such, AUC ROC is robust to highly imbalanced data distributions (as is seen here) (Jeni et al., 2013). A model with an AUC of 0.5 performs at chance, and a model with an AUC of 1.0 performs perfectly.

The question of how well a student would perform on the SAT was treated as a numerical regression problem – i.e. we attempted to predict the student's actual score on the combined mathematics and verbal SAT, ranging from 400 to 1600. We attempted to predict SAT score using linear regression, a standard and straightforward regression algorithm, used in many papers.

The core metric used for assessing the quality of the model predicting SAT score was Spearman's ρ, also referred to as Spearman's rank correlation coefficient (Kendall, 1949). Spearman's ρ indicates the degree to which, when one variable goes up, the other variable goes up as well. In this case, it indicates the degree to which, when our predictive model says a student's SAT score should be high, it is actually high, and when the predictive model says a student's SAT score should be low, it is actually low. Spearman's ρ is preferred to the more common Pearson correlation for situations – like the current one – where one or more of the variables does not follow a normal (bell-curve) distribution. A model with a Spearman's ρ of 0.0 performs at chance, and a mode with a Spearman's ρ of 1.0 performs perfectly.

In both cases, we used the scikit-learn machine learning software within the Python programming language to run these algorithms. Missing values in the data were replaced with the most frequent value in the data, a common practice that has some limitations when conducting statistical significance testing, but which are less relevant in the machine learning/data mining context. In order to control for the large number of features distilled from the data, we selected which features to input into our algorithms using forward selection, where the feature that most improves model goodness is added repeatedly until adding additional features no longer improves model goodness.

Given the significant class imbalance within the graduation data (around three times as many students who graduated than students who did not graduate), we used re-sampling to adjust our training sets for graduation, a standard practice for developing this type of model with imbalanced data. Re-sampling was only used on the training sets; all calculation of model quality took place using unmodified test sets, as discussed in the next paragraph.

Each algorithm was evaluated using 10-fold cross-validation (Efron & Gong, 1983). In this process, students are split randomly into 10 groups. Then, for each possible combination, a model is developed using data from nine groups of students (the “training set”) before being tested on the tenth “held out” group of students. By cross-validating, we can assess how well our models can be expected to function for entirely new students drawn from the same population as our sample.

Results

How effective are predictive analytics models when used within-population or across-population?

The goodness of the best models predicting graduation are shown in Table 1. It was possible to predict eventual graduation from Junior year data both for military-connected and non-military-connected students. The model predicting graduation for military-connected students achieved an AUC of 0.70, indicating that it could distinguish a student who would eventually graduate from a student who would not graduate 70% of the time, for entirely new students. The model predicting graduation for non-military-connected students achieved an AUC of 0.71, indicating that it could distinguish a student who would eventually graduate from a student who would not graduate 71% of the time, for entirely new students. As such, both models were reasonably successful at predicting overall graduation when applied to new students within their population. AUC values in the mid 0.70s are used in medical decision-making with major real-world impact, such as the choice of which anti-retroviral therapy to use for HIV patients (e.g. Revell et al., 2013). As such, while the graduation prediction models presented here are imperfect, they are at a level of quality where they can be used for basic research and intervention, given appropriate caution.

When applied across-population – i.e. a model is developed using data from military-connected students and applied to data from non-military-connected students, or vice-versa – the models remained above chance, but degraded about halfway from their original performance to chance. Specifically, applying the military-connected model to non-military-connected students resulted in an AUC of 0.60, and applying the non-military-connected model to military-connected students resulted in an AUC of 0.60 as well. In other words, when a model was applied to a student from the other population, it could only identify whether they would graduate 60% of the time – halfway between the models' same-population (but new student) performance of 70%/71% and chance performance of 50%. School districts and researchers that use predictive analytics or indicator-based approaches of student at-risk status tend not to pay attention to whether a student is military-connected. If a model is developed on a population that is predominantly not military-connected, it is likely to function substantially more poorly for military-connected students, hampering attempts to support that student. As such, our results lead us to **recommend that at-risk prediction models explicitly consider whether a student is military-connected, and ideally develop a separate prediction model and set of indicators for military connected students.**

Table 1

The performance of the algorithms predicting whether a student will graduate from high school. In all cases, predictions are based on students not used in model development -- student-level cross-validation is used within-population to achieve this. AUC ROC statistic is provided in all cases.

Model developed with data from	Model tested on military-connected students	Model tested on non-military-connected students
Military-connected	0.70	0.60
Non-military-connected	0.60	0.71
Chance	0.50	0.50

The goodness (quality) of the best models predicting the SAT are shown in Table 2. It was possible to predict SAT score both for military-connected and non-military-connected students. The model predicting SAT for military-connected students achieved a Spearman's ρ of 0.53, indicating good ability to predict the SAT; after controlling for non-normality in the data, our model can predict 28.1% of students' variance in eventual SAT scores (ρ^2). This is a fairly large proportion of predictive power, considering the large degree of prediction of the SAT that comes simply from factors such as personality variables, ethnicity/race, birth language, and parent socio-economic status (Noftle & Robins, 2007; Dixon-Roman et al., 2013). The model predicting SAT score for non-military-connected students achieved a Spearman's ρ of 0.52, indicating good ability to predict the SAT; after controlling for non-normality in the data, our model can predict 27.0% of students' variance in eventual SAT scores (ρ^2). As such, both models were reasonably successful at predicting SAT score when applied to new students within their population.

When applied across-population – i.e. a model is developed using data from military-connected students and applied to data from non-military-connected students, or vice-versa – the models remained reasonably successful, but degraded by about a third from their original performance to chance. Specifically, applying the military-connected model to non-military-connected students resulted in a ρ of 0.42 (ρ^2= 17.6%), and applying the non-military-connected model to military-connected students resulted in a ρ of 0.44 (ρ^2= 19.4%). In other words, when a model was applied to a student from the other population, it could only capture about 2/3 as much of the variance in eventual SAT score. This degree of degradation of performance is lower than the degree of degradation for predicting graduation, but is nonetheless substantial – again, leading us to **recommend that**

SAT score prediction models explicitly consider whether a student is military-connected, and ideally develop a separate prediction model and set of indicators for military connected students.

Table 2

The performance of the algorithms predicting a student's SAT score. In all cases, predictions are based on students not used in model development -- student-level cross-validation is used within-population to achieve this. Spearman ρ statistic is provided in all cases, with ρ^2 in parentheses.

Model developed with data from	Model tested on military-connected students	Model tested on non-military-connected students
Military-connected	0.53 (0.28)	0.42 (0.18)
Non-military-connected	0.44 (0.19)	0.52 (0.27)
Chance	0.00 (0.00)	0.00 (0.00)

Understanding which features are important to prediction: Graduation

Models developed using data mining are notoriously difficult to interpret; even relatively interpretable models such as logistic regression and linear regression involve understanding the interrelationships of several (in this case 10-13) variables, which are themselves intercorrelated. Baker (2017) provides an example of how this type of interpretation is highly challenging, even for linear models consisting of only two correlated variables.

As such, we will focus on understanding which features are important to prediction in a different way – considering the features' individual relationships with graduation. To do this, we create a set of single-feature models – logistic regression models that only contain one feature apiece. We can then compare the single-feature models between the two populations, to see where the relationships differ considerably. Note that no single predictor will have a particularly high impact when considered individually, but studying the predictors individually helps us better understand the differences between populations. Note that in evaluating the models' quality, we used cross-validation, which repeatedly splits a data set into subsets used to develop the model and subsets used to evaluate the model; here we are reporting results based on data from all students in the sample, as the best estimate of the overall relationship between each feature and graduation.

When we compare the single-feature models, we find several interesting differences. **Different factors are indicative of at-risk status for military-connected and non-military connected students.** For example, students who are placed into DAEP – Disciplinary Alternative Education Programs – are substantially less likely to graduate high school if they are non-military connected; a similar effect is not seen for military-connected students. Take two students who were both placed into DAEP; one military-connected, the other non-military-connected, but identical in all other features. If the military-connected student had a baseline 75% chance of graduating, the non-military-connected student would have only a 45% chance of graduating.

So, too, the implications of possession of a non-illegal knife are different for military-connected and non-military connected students. A student caught with a knife is much less at-risk if they are military-connected. Take two students who were both disciplined for possession of a knife; one military-connected, the other non-military-connected, but identical in all other features. If the military-connected student had a baseline 75% chance of graduating, the non-military-connected student would have only a 49% chance of graduating.

Similar patterns are also seen for possession of controlled substances and disruptive behavior. These behaviors are far more indicative of risk for non-military students than military students. There are many possible explanations for these differences, but one possible hypothesis is that military parents may respond very differently than non-military parents when their child is placed in DAEP, disruptive, or caught with controlled substances. In addition, the possession of a non-illegal knife may have different implications for a military-connected student than a non-military connected student.

Other factors are more indicative of risk for military-connected students than non-military connected students. For example, a military-connected student who is disciplined for refusing to work is much more at risk than a non-military connected student, perhaps because the refusal to work is more indicative of stress or emotional problems. Take two students who were both disciplined for refusal to work; one military-connected, the other non-military-connected, but identical in all other features. If the non-military-connected student had a baseline 75% chance of graduating, the military-connected student would have only a 48% chance of graduating. A related pattern is seen for inappropriate drawing and writing and for "inappropriate exposure" – these behaviors are much more serious indicators of at-risk status for military-connected students than non-military-connected students.

Another highly predictive feature is one that might be disregarded for non-military-connected students: attending a funeral. Perhaps due to military-connected students' unique circumstances, funerals are a considerably stronger indicator of at-risk status for military-connected students than other students. If a non-military-connected student who attended one or more funerals had a baseline

75% chance of graduating, the military-connected student would have only a 57% chance of graduating.

Table 3 shows a selection of some of the features important to predicting graduation among military-connected and non-military-connected students, focusing on particular on features that differentiate these two populations.

Table 3

The relationship between selected single features and graduation. Numbers shown are in terms of relative probability of graduation for two students – one military-connected, the other non-military-connected – who are identical in all variables except this variable. (Note that relative baseline risk is set to 75% to facilitate comparison; any baseline could be chosen).

Feature	Military-connected graduation probability	Non-military-connected graduation probability
Student placed in DAEP	75%	45%
Possession non-illegal knife	75%	49%
Possession of Controlled Substances	75%	46%
Disruptive behavior	75%	46%
Refusal to work	48%	75%

Inappropriate Drawing and Writing	49%	75%
Inappropriate Exposure	59%	75%
Attending Funeral	57%	75%

Understanding which features are important to prediction: SAT

As with our work to interpret the relationships between student data features and graduation, we interpret the relationships between the features and the SAT by creating a set of single-feature models – linear regression models that only contain one feature apiece. We can then compare the single-feature models between the two populations, to see where the relationships differ considerably. As above, no single predictor will have a particularly high impact when considered individually, but studying the predictors individually helps us better understand the differences between populations.

When we compare the single-feature models, we find several interesting differences. For example, while final grade in science class is predictive of better SAT scores for both military-connected and non-military connected students, the effect is over twice as large for non-military connected students. A non-military connected student who gets an A in science class is likely to perform 145 points better on the total SAT than a non-military-connected student who gets a C in science class. By contrast, a military-connected student who gets an A in science class is only likely to perform 56 points better on the total SAT than a student who gets a C in science class – still a difference, but a substantially smaller one. Similar findings are also seen for whether students took fine arts classes and foreign language classes – they are more predictive for non-military-connected students than military-connected students.

By contrast – and somewhat surprisingly –disciplinary data was more predictive of poorer SAT scores for military-connected students than other features, primarily relatively minor infractions. Disciplinary data were also more predictive of SAT scores for military-connected students than non-military-connected students. For example, each instance of a hall pass violation was associated with performing 137 points worse on the total SAT for military-connected students; by contrast, hall pass violations had around half as strong a relationship to total SAT (71 points worse) for non-military-connected students. Relatedly, each instance of being disciplined for failing to comply with teacher instructions was associated with performing 72 points worse on the total SAT for military-connected students; there was no relationship between failure to comply and total SAT for non-military connected students. Similar patterns were also seen for being sent to the office by the teacher and disruptive behavior – in these cases, while these behaviors were associated with lower SAT scores for military-connected students, they were unexpectedly associated with higher SAT scores for non-military-connected students (albeit to a much lower degree).

Table 4 shows a selection of some of the features important to predicting SAT among military-connected and non-military-connected students, focusing on particular on features that differentiate these two populations.

Table 4

The relationship between selected single features and SAT. Numbers shown are in terms of expected change in SAT score for two students – one high in the feature, the other low in the feature -- who are identical in all variables except this variable.

Feature	Expected difference in SAT score: military-connected	Expected difference in SAT score: non-military-connected
A in Science class versus C in Science class	+56	+145
Took Fine Arts class versus Didn't Take Fine Arts	-4	+34
Took Foreign Language versus Didn't Take F. Lang	-2	+36

1 Instance of Failure to Comply versus 0 Instances of Failure to Comply	-72	+/- 0
1 Instance of Hall Pass Violations versus 0 Instances of Hall Pass Violations	-137	-71
1 Instance of Sent to Office versus 0 Instances of Sent to Office	-78	+17
1 Instance of Disruptive Behavior versus 0 Instances of Disruptive Behavior	-109	+50

Conclusions

Within this paper, we have examined whether the common practice of using a single at-risk prediction model for both military-connected and non-military-connected students leads to valid prediction. We find evidence for substantial degradation in model performance; while the models remain above chance when applied across population, they produce substantially poorer performance than a model specialized to military-connected students. In particular, we hypothesize that this finding is because many factors differ in their meaning between military-connected and non-military-connected groups of students. A non-military-connected student who brings a non-illegal knife to school is far more at-risk than a military-connected student who brings such a knife to school; these students may have brought these knives to school for very different reasons. Non-military-connected students are not much more at-risk if they miss school to attend a funeral; military-connected students who attend funerals are at much higher risk. Even when it is not a family member who passed away, attending a funeral may be a reminder of a loved parent who is themselves in danger during a deployment. Finally, suspensions and disciplinary alternative education programs are much higher indicators of risk for non-military-connected students, possibly due to different parental responses to these circumstances.

In interpreting these factors, however, it is important to recognize that the models presented here give no evidence with regards to causality. We do not know precisely why a military-connected student's refusal to work is associated with lower probability of graduation. The refusal to work probably does not ***cause*** dropout, so much as both dropout and a refusal to work are the result of some other problem in the student's life. As such, if a seemingly unusual predictor such as this one is particularly relevant for a given student, it may provide an opportunity for further probing and problem-solving on the part of school personnel.

We believe that the next key step for this work is to develop and test these models for other school districts as well, to see if similar patterns of results are obtained, in terms of the cross-applicability of models developed for military-connected and non-military-connected students. We can also use data from other school districts to determine whether an at-risk prediction model developed for military-connected students in one district is applicable to military-connected students in other districts as well.

An additional valuable next step with these models is to deploy them in a school on an ongoing basis, and see whether the model predictions can form the basis of meaningful intervention. We intend first to develop reports, building on earlier work that attempted to derive general design principles for creating reports on student at-risk status (Ocumpaugh et al., 2017). We will use co-design (Penuel et al., 2007) to develop these reports, working with the users (school personnel such as principals) to determine how best to communicate what the model has determined for use in intervention. We will deploy and iteratively enhance these reports in partnership with school personnel, seeing how to improve the information available and seeing what practices work most effectively with these reports. Beyond simply providing an end-of-semester risk estimate, this model can also be used to identify key situations, previously less focused on by school personnel, where probing to determine what is happening may be particularly helpful, as discussed earlier in this section. Finally, we intend to conduct a study to evaluate whether our data-driven intervention approach can be successful at concretely increasing graduation rates and SAT scores.

References

Allensworth, E. (2013). The use of ninth-grade early warning indicators to improve Chicago schools. *Journal of Education for Students Placed at Risk (JESPAR)*, *18*(1), 68–83.

Allensworth, E. M., & Easton, J. Q. (2007). *What matters for staying on-track and graduating in Chicago public highs schools: A close look at course grades, failures, and attendance in the freshman year.* Chicago, IL: University of Chicago Consortium on Chicago School Research.

Allensworth, Gwynne, J., Moore, P., & de la Torre, M. (2014). *Looking forward to high school and college: Middle grade indicators of readiness in Chicago public schools*. Chicago, IL: University of Chicago Consortium on Chicago School Research.

Baker, R. (2017). *Big data and education* [MOOC]. 3rd ed. Philadelphia, PA: University of Pennsylvania.

Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 253–272). Cambridge handbooks in psychology. New York, NY: Cambridge University Press.

Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, *42*(4), 223-235.

Baltimore Education Research Consortium. (2011). *Destination graduation: Sixth grade early warning indicators for Baltimore city schools.* Baltimore, MD: Baltimore Education Research Consortium.

Bowers, A. J., Krumm, A., Feng, M., & Podkul, T. (2016). *Building a data analytics partnership to inform school leadership evidence-based improvement cycles.* Paper presented at the annual meeting of the American Educational Research Association. Washington, D.C.

Bowers, A. J., Sprott, R., & Taff, S. A. (2013). Do we know who will drop out?: A review of the predictors of dropping out of high school: Precision, sensitivity, and specificity. *The High School Journal*, *96*(2), 77–100.

Carl, B., Richardson, J. T., Cheng, E., Kim, H., & Meyer, R. H. (2013). Theory and application of early warning systems for high school and beyond. *Journal of Education for Students Placed at Risk (JESPAR)*, *18*(1), 29–49.

Chandra, A., Martin, L. T., Hawkins, S. A., & Richardson, A. (2010). The impact of parental deployment on child social and emotional functioning: Perspectives of school staff. *Journal of Adolescent Health, 46*(3), 218–223

De Pedro, K., Astor, R.A., Benbenishty, R, Estrada, J.N., Smith, G.R., & Esqueda, M.C. (2011). The children of military service members: Challenges, resources, and future educational research. *Review of Educational Research, 81*(4), 566-618.

Dixon-Román, E. J., Everson, H. T., & McArdle, J. J. (2013). Race, poverty and SAT scores: Modeling the influences of family income on black and white high school students' SAT performance. *Teachers College Record*, *115*(4), 1-33

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, *37*(1), 36–48.

Hanley, J., and McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology, 143*, 29–36.

Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). Facing imbalanced data—Recommendations for the use of performance metrics. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction* (pp. 245–251).

Kemple, J. J., Segeritz, M. D., & Stephenson, N. (2013). Building on-track indicators for high school graduation and college readiness: Evidence from New York City. *Journal of Education for Students Placed at Risk (JESPAR)*, *18*(1), 7–28.

Kendall, M. G. (1949). Rank and product-moment correlation. *Biometrika*, 177-193.

Kieffer, M. J., & Marinell, W. H. (2012). Navigating the middle grades: Evidence from New York City. New York, NY: The Research Alliance.

Mmari, K., Roche, K., Sudhinaraset, M., & Blum, R. (2009). When a parent goes off to war: Exploring the issues faced by adolescents and their families. *Youth and Society, 40*(4), 455–475.

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: big five correlates of GPA and SAT scores. *Journal of personality and social psychology*, *93*(1), 116

Ocumpaugh, J., Baker, R., Slater, S., San Pedro, M.O., Heffernan, N, Heffernan, C., Hawn, A. (2017). Guidance counselor reports of the ASSISTments college prediction model (ACPM). *Proceedings of the International Conference on Learning Analytics and Knowledge*, 479–488.

Palmer, C. (2008). A theory of risk and resilience factors in military families. *Military Psychology*, *20*(3), 205-217.

Penuel, W., Roschelle, J., Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. *Research & Practice in Technology Enhanced Learning, 2*(1), 51–74.

Revell, A. D., Wang, D., Wood, R., Morrow, C., Tempelman, H., Hamers, R. L., Alvarez-Uria, G., Streinu-Cercel, A., Ene, L., Wensing, A.M.J., DeWolf, F., Nelson, M., Montaner, J.S., Lane, H.C., Larder, B.A. (2013). Computational models can predict response to HIV therapy without a genotype and may reduce treatment failure in different resource-limited settings. *Journal of Antimicrobial Chemotherapy, 68* (6), 1406-1414.

Tobin, T. J., & Sugai, G. M. (1999). Using sixth-grade school records to predict school violence, chronic discipline problems, and high school outcomes. *Journal of Emotional and Behavioral Disorders*, *7*(1), 40-53.

Wiltz, J., & Slate, J. R. (2016). Differences in dropout rates by ethnicity/race of middle school students: A multi-year analysis. *Global Journal of Human-Social Science Research, 16*(8).