

EXAMINING DIVERSITY IN DEVELOPMENTAL TRAJECTORIES OF CADETS' PERFORMANCE AND CHARACTER AT THE UNITED STATES MILITARY ACADEMY

Hillary S. Schaefer, Kristina Schmid Callina, Jeremiah Powers
Tufts University

Gerald Kobylski
United States Military Academy

Diane Ryan and Richard M. Lerner
Tufts University

A developmental model linking professionalism and character growth is implemented at the U.S. Military Academy at West Point (USMA) through its programs and policies. The current study used growth-modeling techniques to examine developmental relationships between metrics of professional development and performance outcomes at USMA and assessed whether race, gender, and athlete status variables moderated the resulting patterns. Data were collected from cadets in the classes of 2017 and 2018, in their first through third years ($N = 2,066$). Metrics included grade point average (GPA) for the academic, military, and physical programs, and the periodic development review (PDR), a 23-item instrument developed by USMA that evaluates cadets on various professional values, including character attributes. Results indicated that academic and military program GPAs, PDRs, and honor violations were interrelated, with important group differences in GPAs and PDRs by race and between National Collegiate Athletic Association athletes and nonathletes. Models suggested resilience across the corps of cadets from the first to third year; however, group differences favored White and nonathlete cadets in GPAs and PDRs, including self-ratings, even when SAT was accounted for. Implications for junior officer professional development, as well as for professional ratings systems are discussed.

The educational model at the U.S. Military Academy (USMA) seeks to integrate character development within its approach to leadership

training, performance standards, and internalization of Army values. The USMA model of professional preparation thus emphasizes a

• Correspondence concerning this article should be addressed to: Hillary Schaefer, hillary.schaefer@tufts.edu

Journal of Character Education, Volume 17(1), 2021, pp. 59–80
Copyright © 2021 Information Age Publishing, Inc.

ISSN 1543-1223
All rights of reproduction in any form reserved.

combination of traditional academics, learning skills relevant to the profession-of-arms, and understanding the profession's ethical standards and conceptions of being a leader of character (Callina et al., 2018; Colby & Sullivan, 2008). Metrics associated with this integrated developmental model at USMA include professional evaluations that provide feedback on values important to the system, such as character and leadership attributes, as well as grades across the three training programs: academic, military, and physical (USMA, 2018). Program grades are weighted and comprise the metric that determines USMA class rank, which in turn has implications for cadets' subsequent career trajectories in the U.S. Army. The focus on program grades as the key determinant in cadet promotion signifies the importance USMA places on competitive spirit, maintaining standards of performance, and demonstrating excellence. In contrast, cadet professional development metrics are used for evaluation and advising. These character-relevant items explicitly outline virtues important to the Army (U.S. Department of the Army, 2012), such as leadership, professionalism, and discipline.

Character attributes such as grit and resilience are correlates of success at USMA as measured by individual components of its evaluation system, such as grade point average (GPA) and retention (Bartone et al., 2002; Maddi et al., 2017). The professional development model championed at USMA includes these empirically derived and validated concepts; however, as with many professional evaluation systems, the USMA model also incorporates other values that are critical to the environment but without such longstanding empirical background, such as the ability to build trust in subordinates, and other character-adjacent or character-modifying attributes such as leadership and empathy. The present study sought to expand on prior work connecting individual character attributes to USMA performance by examining multiple components of the larger professional preparation model over the course of USMA officership

development. Accordingly, we sought to examine trajectories of scores on professionally-relevant virtues and performance metrics, and determine whether these metrics were related to one another longitudinally. We also investigated whether cadet development as measured by these institutionally-derived metrics could be predicted by individual and group differences, such as SAT scores, athlete status, or demographic characteristics. Information regarding the relations among performance, character, and professional growth at USMA and potentially diverse developmental pathways might illuminate strengths of USMA's professional preparation model, and also possibly recommend opportunities for improvement.

Character and Performance Assessment at USMA

As noted above, performance at USMA is evaluated by scores in the academic, military, and physical development programs. Assessments of character attributes such as grit, optimism, gratitude, and so on have proliferated in recent years within psychology, education, and related fields, but the nature of the development of character is still under-studied (see Callina et al., 2017, and Lerner & Callina, 2014, for brief reviews of this history). Attributes such as grit, bravery, zest, fairness, honesty, persistence, optimism, leadership, self-regulation, and teamwork appear to be critical for soldier performance in the U.S. Army generally (Eskreis-Winkler et al., 2014; Von Culin et al., 2014), but how these emerge or change developmentally is still unclear. These and similar attributes predict retention and academic achievement at USMA (Duckworth et al., 2007; Kelly et al., 2014), but such characteristics are only useful for repeated assessment as part of USMA's character and professional development strategy if they are responsive to developmental processes such as intervention or maturation. In other words, if grit is an important, but stable and context-independent characteristic, it is a good

metric for selection of USMA cadets but may not be the most efficient focus of assessment or advising during their tenure, as compared to a characteristic that is very malleable within the 47-month-long developmental window for USMA cadets (Callina et al., 2017, 2018). The nature of developmental pathways of professional growth and performance, and the relations between these pathways between and within individuals, is a yet understudied aspect of the science of character (Lerner, 2018).

Some character assessment systems have empirical origins (e.g., the VIA Inventory of Strengths; Peterson & Seligman, 2004), whereas others are developed “in house,” and represent the values of the organization and of the preceding leaders who have shaped the instruments. Concerns about character and moral attributes, and how to evaluate them, have flourished in recent years in business (Cohen & Morse, 2014; Van Iddekinge et al., 2012); education, including higher education (Berkowitz, 2012; Hershberg et al., 2016; Jeynes, 2017); and in militaries across the world (Boe et al., 2015; Callina et al., 2017; Currier et al., 2015; Yu, 2018). Current theory suggests that such character assessments that are specific to a context, such as the military, are more appropriate for understanding character as compared to more global assessments that might be equally applicable across many environments (Callina et al., 2017).

As such, the professional development tools considered here highlight the context-specific character and personal virtues that are desired by USMA. USMA defines character as having five key facets: moral, civic, performance, social, and leadership (USMA, 2018). Their metrics and programs align with this “in house” model and include components somewhat afield of empirically derived measures of character, such as adherence to the mission and physical fitness. However, demonstrating these crucial attributes in the specific manners emphasized in this specific context reflects the Aristotelian concept of *phronesis*, or enacting the appropriate characteristics in the appropriate time and place.

Whereas physical fitness or mission adherence might be irrelevant in another context, display of these attributes demonstrates character and professionalism for USMA cadets. Understanding how this instance of *phronesis* is enacted and developed has the potential to illuminate as much about professional development generally as it does about these specific character attributes. Therefore, despite the specificity of USMA’s metrics with respect to change across the 47-month educational experience, understanding its system of evaluation, individual differences, and relations to other important outcomes can help any institution invested in professional growth and character development to create and/or tailor a system specific to their own contextually derived values.

In the military, chain-of-command ratings of professional virtues are a primary component of the military’s promotion system (Moore & Trout, 1978). This system is especially relevant for early- and midcareer servicemembers (Bowman & Mehay, 1999; Harris, 2009), despite a number of criticisms due to various biases in the promotion system (Butler, 1999; Chapman, 2006; Hopkins & Williams, 2013; Kane, 2011). The periodic development review (PDR) was developed by USMA and is based on the evaluation system used with commissioned and noncommissioned Army officers. It contains 23 items that were derived from Army doctrine (U.S. Department of the Army, 2012), relevant to cadet character and professional development, and aligns with their model of the five character facets. Completed by instructors, supervising officers, other cadets, and cadets themselves (the self-version), PDRs capture observations of cadet character and professional development throughout their 47 months at USMA and from multiple viewpoints. Using the PDR, cadets acquire experience as both the rater and ratee in settings across the military, academic, and physical program at USMA.

Analysis of institutionally derived professional evaluation tools such as the PDR is

important if researchers seek to understand what is being specifically reinforced by the larger system and who or what is considered “successful” in the specific context. As noted above, understanding the approaches to evaluation used in any one specific system can provide guidance for educational and professional systems seeking to tailor assessments to their own values. Post-hoc evaluations of such systems often determine variables associated with “success” or “failure” on these instruments (e.g., Hanser & Oguz, 2015; Mueller & Mazur, 1996), without first establishing validity of the measure or understanding its underlying factor structure. To address this issue, the current study first explored the factor structure of USMA’s professional development instrument, the PDR, assessed its associations with program grades and behavioral outcomes (honor violations), and examined evidence of individual and group differences in these trajectories of performance and professional development.

Potential Diversity in Performance and Character Scores

Issues of race, gender, and diversity are important to the U.S. Army. USMA’s character development mission statement explicitly outlines the importance of diversity, leadership of diverse teams, and inclusiveness (USMA, 2018). Despite this intent, and like the greater social context in which the military is embedded, servicemembers of color face a variety of challenges. Race (Burk & Espinoza, 2012; Hopkins & Williams, 2013) and gender (Baldwin, 1996) differences in promotion rates are a pervasive problem in the military, and problems of sexual assault/harassment remain in the Army generally (Street et al., 2008; Turchik & Wilson, 2010), and USMA specifically (Arbeit, 2016). Therefore, the specific ways in which group differences might be associated with developmental trajectories of cadets’ professional growth and performance is important to the current investigation.

A particular concern with performance ratings systems, especially those with more abstract targets such as character virtues, is that of implicit bias. Implicit bias refers to preferences or tendencies that occur outside of conscious awareness (e.g., Lai et al., 2013), and is often measured through preferential selection or association of images, words, or ideas. These implicit attitudes can be manifested as behavior via differences in social interactions (Arkes & Tetlock, 2004; Dovidio et al., 2002; Greenwald et al., 2009) and judgments of others (Amodio & Devine, 2006), especially when motivations to control prejudice are low (Crandall & Eshleman, 2003). Although critiques of the implicit bias literature have noted that this construct may better reflect cultural associations than personal animosity (Arkes & Tetlock, 2004; Walker et al., 2015), both associations and personal beliefs are relevant to metrics of performance, character, and professional development.

The phenomenon of implicit associations may impact professional development and performance scores through several avenues, including: (1) Raters may score ratees based on stereotypes rather than actual performance (Gralewski & Karwowski, 2013; Pigott & Cowen, 2000; Tenenbaum & Ruck, 2007); (2) Raters may change their interactions with ratees in a way that alters their perception of a ratee’s performance or professionalism, creating a self-fulfilling prophecy effect (Chen & Bargh, 1997; Kunda & Spencer, 2003); (3) Raters may experience confirmation bias, wherein stereotype-congruent behaviors are remembered more clearly than stereotype-incongruent ones (Darley & Gross, 1983); or (4) Ratees holding even implicit stereotypes about their own group may underperform due to aversion to potentially enacting stereotype-consistent behaviors, and possibly resulting in decreased self-esteem or reduced goal-setting as well as increased chronic stress (Robins & Pals, 2002). Considerable evidence exists for each of these effects in both performance outcomes as well as in ratings systems and related assessments, such as performance

evaluations or beliefs about abilities (Grunspan et al., 2016; Riegle-Crumb & Humphries, 2012). These effects are particularly evident for performance evaluations and career trajectories for women in stereotype-inconsistent jobs (Lyness & Heilman, 2006), including the military (Baldwin, 1996; Boyce & Herd, 2003; Prividera & Howard, 2006). Furthermore, prejudicial beliefs, particularly against women, shape servicemembers' beliefs about others' leadership and competence (Boldry et al., 2001), are greater in military versus comparable civilian contexts (Matthews et al., 2009), and appear to increase with time of service (Biernat, 2003; Boyce & Herd, 2003).

Whereas the current investigation is unable to determine precisely how metrics may be biased for or against a particular group, it is important to understand whether specific developmental trajectories are influenced by race or gender, and whether that influence resembles bias that might be impugning cadets' achievement, and perhaps ultimately, their military careers. We hypothesized that PDR scores would relate to character- and professionalism-relevant behavioral outcomes, and also that demographic biases would remain across both professional development metrics and program grades. Importantly, ratings and performance evaluations may have differential impacts on career-relevant outcomes depending on demographic group (Greenhaus et al., 1990), and so we further expected that the relationship between PDR scores, program grades, and behavioral outcomes might vary according to group membership.

The Current Investigation

In the present study, we employed growth-based trajectory models to examine how PDR scores by various sources (instructor, peer cadet, and self) related to program grades across the three USMA training pillars, and how these trajectories related to demographic variables including race and gender, SAT scores, and honor violations. We also

included National Collegiate Athletic Association athlete status as a demographic category, given the status of athletes as a group that receives additional privileges but also has extraneous demands; this difference exists at USMA but is also true of student athletes in higher education generally (Adler & Adler, 1985; Gaston-Gayles, 2004).

The analysis strategy used structural equation modeling techniques, in particular latent growth modeling (Figure 1). Growth modeling techniques use repeated measurement to estimate initial levels (intercepts) and growth rates (slopes), examine the interrelationships (covariances) between slopes and intercepts of different measures, and test whether group membership or other time-invariant variables are related to these relationships. Using this technique, the central hypothesis was that professional growth as measured by the PDR, especially PDRs from instructors, would relate to growth in program grades. We specifically hypothesized that program grades and PDRs would be related in terms of their slopes and intercepts. However, we also expected that race and gender would moderate the ways in which key relationships between professional growth metrics and outcomes operated, in line with previous research (McCaslin, 2016; Sellers, 1992; J. L. Williams & Deutsch, 2016).

As noted above, PDRs for each cadet are completed by instructors, peers, and cadets themselves. Self-ratings are a frequent component of organizational and performance evaluations, both as a discussion tool and as an opportunity for the ratee to self-advocate. Despite such uses, ratings of the self frequently suffer from several biases, including social desirability, insight, and self-deception (Donaldson & Grant-Vallone, 2002; Van de Mortel, 2008). Generally, individuals who have the lowest abilities are the least accurate in their own self-assessments (Hall & Raimi, 2018; Kruger & Dunning, 1999), a finding that holds in military leadership ratings as well (Bass & Yammarino, 1991; Pazy & Oron, 2001). Given this finding, we hypothesized that self-completed PDRs would be the least accurate of all

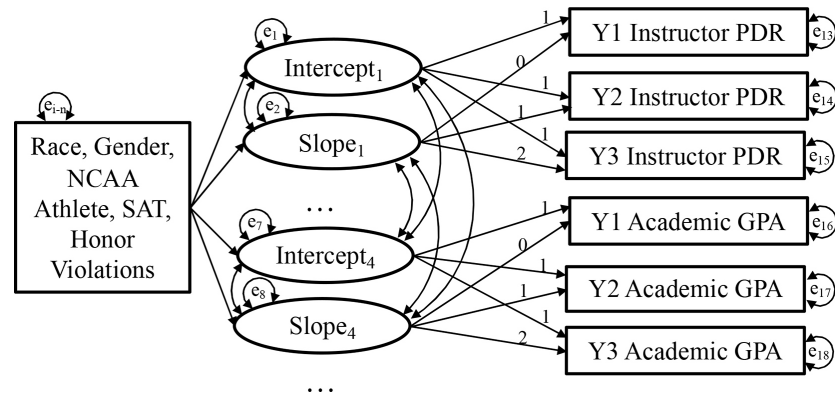


FIGURE 1

Example growth model depicting the relations among PDR, GPA, their slopes and intercepts, and covariates. The latent intercept is fixed to 1 at each time point, and the latent slope is fixed to 0, 1, and 2 at each time point respectively in order to estimate a linear growth term. For clarity, only one PDR rater and GPA type are shown; in the full model 6 slopes and 6 intercepts are modeled (Instructor, cadet, and self PDR, academic, military, and physical GPA). Similarly, race, gender, National Collegiate Athletic Association athlete status, SAT, and honor violations were input as separate time-invariant predictors but shown together for clarity; S = Slope, I = Intercept, Y1 = freshman year, Y2 = sophomore year, Y3 = junior year.

versions of the measure in predicting professional development outcomes.

Finally, we were interested in understanding how these trajectories of professional development and performance metrics might relate to objective behavioral outcomes at USMA. Violations of USMA's honor code ("A cadet will not lie, cheat, or steal, or tolerate those who do") represent a relatively rare, but behaviorally important, lapse in ethics, conduct, and judgment. The system for identifying, adjudicating, and remediating honor code violations is an important component of training for USMA cadets and can be considered the minimally acceptable standard of ethics and conduct for the profession of arms (Cushen et al., 2012). Thus, understanding how the metrics of professional development and performance relate to students who are not meeting this standard is an important and commonly used outcome for understanding the behavioral manifestations of character, and we hypothesized these lapses in judgment would be associated with lower scores on profes-

sional development metrics and also lower program grades (Carrell et al., 2009; Samuels & Casebeer, 2005).

METHOD

Participants and Data Sources

Data from 2,066 cadets from the first 3 years of the graduating classes of 2017 to 2018 were included in growth models (18% female; 69% White, 10.7% Black, 7.7% Asian, 9.7% Hispanic, 2.8% Other; 18.4% NCAA-recruited athlete).¹ As noted below, PDRs from cadets' final year at USMA show very little variation due to the scale's construction, and so data from the fourth-year cadets were excluded from these analyses. Only cadets completing the first 3 years, and with valid PDR and GPA data, were included. The incoming classes for these two cohorts were initially a combined total of 2,420 students for the graduating classes of 2017 and 2018; by their third year a

total of 354 had separated, resigned, were on administrative leave, or otherwise had missing data. To conduct the exploratory principal component analysis of PDR items, we used data from 4,903 cadets from the classes of 2017 to 2020. These data included PDRs completed by instructors, cadet peers, and a self-rating for each semester of the cadet's attendance (see below for details on PDR administration). International exchange cadets were excluded from all analyses as they are generally in attendance for one or two semesters, may not be rated by all sources, and thus have an insufficient sample of PDRs for analysis. All data were obtained from USMA's data warehouse and coded by an anonymous ID in accordance with institutional review board procedures.

Measures

PDR Procedures and Format. The periodic development review (PDR) is a 23-item instrument developed by the USMA. The instrument used Army doctrine (i.e., the specific character virtues and professional skills required for excellence as a commissioned U.S. Army officer) to generate items intended to evaluate cadets on a range of character and professional facets, and to provide training in chain-of-command ratings, mentorship, and self-reflection. In short, the items describe the Army professional ethic as defined by doctrine (U.S. Department of the Army, 2012), and the specifics of that format (online interface, item descriptions) have been refined by USMA over time to achieve developmental, professional, and administrative goals. Each semester, cadets complete a PDR on themselves, and receive one from an instructor and from approximately three other cadets. Cadet-completed PDRs are spread among peer year, subordinate, and chain-of-command as appropriate (e.g., freshmen are rated by one other freshman and one chain-of-command, seniors by one other senior and one subordinate). PDRs are distributed and completed through a general online portal used for grades,

class administration, and so on. In-person advising sessions are completed by instructors for instructor-completed PDRs to further the developmental intent of the instrument, whereas cadet-rated PDRs are anonymous. A total of 6,420 different cadets and 1,289 USMA staff (active duty military and civilian) served as raters in the current data sample. For the purpose of analyses, PDRs were averaged across semesters, to give an average score for the year by rating source.

The format of the PDR is presented in Table 1. Items are assessed on a 0-4 scale (0 = *not observed*, 1 = *unsatisfactory*, 2 = *developing*, 3 = *effective*, and 4 = *exceptional*); use of "0" is explicitly discouraged and values of 1 are relatively uncommon, especially after the first year. An overall score is also given but was replaced with the average of the 23 items for the purpose of this study. All 0s were counted as missing values in analyses. The scale for the PDR is intended to be developmental and not age-referential, such that a score of 4 is expected for graduating seniors and scores of 2 are common for freshmen. PDRs from the final year at USMA show very little variance, as most students receive mostly 4s, the highest score; as a result, models were limited to students in the first 3 years. Descriptions are presented for each of the items and 1-4 score levels (see Table 1 for an example). Formal training acclimates new instructors to the system before they complete any PDRs for cadets.

Grade-Point Averages. Each year's cumulative GPA from the three programs (academic, military, and physical) was entered into models. Although a cadet's total GPA is currently computed by weighting the academic GPA 55%, military GPA 30%, and physical GPA 15%, we entered the three program GPAs into the model separately to determine the associations among them, and to understand how PDRs relate to the different program scores.

Each semester's GPA for the three programs is a combination of course grades and, in the case of the military and physical GPA, of other ratings and performance scores. The for-

TABLE 1
Items, Descriptions, and Example of Additional Available Guidance for the PDR

<i>Item Name</i>	<i>Item Description</i>
Army values	Lives, acts, and teaches loyalty, duty, respect, selfless services honor integrity. and personal courage
Empathy	Able to see something from another person's point of view, identifies with and enters into another person's feelings and emotions
Warrior/service ethos	Shares attitudes and beliefs that embody the spirit of the Army profession for soldiers and Army civilians alike.
Discipline	Controls own behavior according to Army values, obeys and enforces orderly practices, does what is morally, legally, and ethically right
Military/professional bearing	Projects a commanding presence and professional image of authority
Physical fitness	Has sound health, strength, and endurance that supports one's emotional health and conceptual abilities under stress
Confidence	Projects self-confidence and certainty, demonstrates composure and poise, calm and collected, possesses self control of emotions
Resilience	Shows a tendency to recovery quickly from setbacks, shock, adversity, stress or injury while maintaining a mission and organizational focus
Mental agility	Flexible of mind, anticipates or adapts to ever-changing conditions, improvises, able to apply multiple perspectives & approaches
Innovation	Able to introduce new ideas based on opportunity or challenging circumstances, original in thoughts & ideas, creative
Expertise	Possesses facts, beliefs, and logical assumptions in relevant areas, technical, tactical, cultural, and geopolitical knowledge
Sound judgment	Assesses situations and draws feasible conclusions, makes sound and timely decisions
Interpersonal tact	Has capacity to understand interactions with others, aware of how others see you and how to interact with them effectively
Leads by example	Provides the example to others, serves as a role model, maintains high standards in all aspects of behavior and character
Leads Others	Motivates, inspires, and influences others to take initiative, work toward a common goal, and accomplish critical tasks and missions
Builds trust	Empowers subordinates, encourages initiative, reinforce accountability and allows open communication
Extends influence	Influences others outside change of command, involves indirect means of influence: diplomacy, negotiation, conflict resolution, and coordination
Communicates	Clearly expresses ideas to ensure understanding, actively listens to others, and employs effective communication techniques
Creates a positive environment	Establishes and maintains positive expectations/attitudes to support effective work behaviors, relationships, and organizations
Prepares self	Conducts self-study, aware of their limitations and strengths, and seek self-development, continues to improve and prepare
Develops others	Encourages and supports others to grow as individuals and teams, prepares others for success, makes the organization versatile and productive
Stewards the profession	Acts to improve the organization beyond their own tenure and supports developmental opportunities for subordinates
Gets results	Consistently produces results, develops and executes plans while providing direction, and clear priorities toward mission accomplishment

TABLE 1
CONTINUED

<i>Example Item Guidance for Discipline</i>				
	<i>Unsatisfactory (Score of 1)</i>	<i>Developing (Score of 2)</i>	<i>Effective (Score of 3)</i>	<i>Exceptional (Score of 4)</i>
Behavioral examples of the category	• Unable to control their behavior and ignores what is right. • Blames others for substandard outcomes. • Verbally critiques or criticizes the importance and role of standards while blaming others when tasks fall below the established standards. • Follows unclear principles and is questionable in word and deed. • Does the wrong thing for no reason and for the wrong goal.	• Inconsistent behavior requiring heavy influence by others to do what is right. • Does not see the need to take responsibility for substandard outcomes. • Takes no verbal stance on the importance and role of standards and feels unable to take action when tasks fall below standards. • Unsure which principles to follow and is sometimes unsure in word and deed. • Does the right thing but is uncertain of the reason or goal.	• Able to control one's own behavior and is compliant to do what is right. • Takes responsibility for substandard outcomes. • Acknowledges the importance and role of standards and takes action when tasks fall below established standards. • Follows clear principles and are honest in word and deed. • Does the right thing for the right reason.	• Behavior is always in accordance with Army values and never waivers from doing what is right. • Takes ownership of all substandard outcomes and seeks improvement. • Reinforces the importance and role of standards throughout the unit and ensures that tasks do not fall below established standards. • A role model for all that follows clear principles and is honest in word and deed. • Does the right thing for the right reason and with the right goal in mind "when no one is looking"

mulas from which the GPA is derived vary somewhat by grade year, and are subsequent to small modifications over time to better capture the mission statements of the program. A full elucidation of each semester's components by program and year is outside the scope of this article, but a few details are noted (see USMA, 2016a, 2016b for more information on program details). The military GPA is composed of military-focused coursework and performance/leadership grades given by the supervising officer. This performance/leadership component is determined by the supervising officer's impressions of a cadet's performance in leadership domains, and may have input from a more senior, supervising cadet. The physical GPA is composed of coursework, fit-

ness testing, and a measure of sportsmanship. Notably, for the first 2 years at USMA the physical program coursework is based mostly on physical performance standards (e.g., the swimming grade), and afterwards has traditional, didactic content (e.g., a course about fitness training schedules) and sport activities taken for fun (e.g., SCUBA or cycling). For the preponderance of students, the physical GPA increases across years.

Scholastic Aptitude Test. SAT and/or ACT scores are required for USMA admission. For students who took only the ACT, their scores were converted to an corresponding SAT scale based on conversion information provided by the College Board, which develops the SAT (Dorans, 2004; Marini et al.,

2016). For students taking both the ACT and SAT, the higher score (actual SAT or converted) was used, which is the typical procedure for USMA's admission process.

SAT scores were used as a covariate predictor in structural equation models estimating the slopes and intercepts of professional growth and performance metrics, meaning that it was a covariate of academic performance. This use complicates the interpretation of the intercept of academic GPA, in that variation associated with a moderate to high correlate of academic performance is removed (Cohn et al., 2004; Young, 2001), a decision contraindicated for some hypotheses (Dennis et al., 2018). The zero-order correlation between freshman year academic GPA and SAT is $r(4790) = 0.64$ ($p < 0.001$) for the current sample. However, SAT is also associated with racial (Eitzen & Purdy, 1986; Fleming, 2002) and socioeconomic (Wells et al., 2016) disparities in educational background, as well as with disparities due to participation in athletics (Sedlacek & Adams-Gaston, 1992). Therefore, including SAT as a covariate allowed us to interpret significant group differences in GPA and professional growth as not associated merely with preexisting disparities or related differences, and instead possibly associated with biases, current educational disparities, and other differences active at USMA.

Honor Violations. The cadet honor code is a foundational system at USMA, and violations of its rules are peer investigated and adjudicated in a court-like system. Either cadets or USMA staff can raise concerns about an honor violation for a cadet. After a conviction for an honor offense, cadets are typically enrolled in an intense mentorship system for 1:1 mentorship and development on topics related to the offense. Example violations include cheating, lying to superiors (e.g., about one's location), and acquiring illegally-downloaded materials. Other types of misconduct, are handled through separate legal channels and were not considered in these analyses. Of the 192 total honor convictions, 58% were for cheating, 38% for lying, and .5% for stealing or related

offenses (e.g., online piracy). Due to the small number of cases the stealing category was excluded for the purpose of this analysis. Due to the sometimes lengthy nature of the investigation, adjudication, and disposition process (which is ultimately decided by USMA's superintendent), a specific year or time frame of the actual offense was not entered into the model. Presence of an honor conviction for cheating or lying was entered into growth models as separate, binary predictors.

RESULTS

Preliminary Analyses

Principal component analyses were conducted on instructor-, cadet-, and self-completed PDRs to determine underlying factor structure of the measure, using the FactoMineR package in R (Le et al., 2008). The 23 items were entered into separate principal components analyses extracting 5 components from the peer, instructor, and self-rated versions of first-year cadets, averaged across the two semesters and two raters in the case of peer report ($n = 4,903$ for each rater type). In each case only a single factor with an eigenvalue > 1.0 was found (first two eigenvalues = 3.72 and 0.37 for instructor, 5.90 and 0.54 for peer, 5.81 and 0.60 for self), likely due to the limited 1–4 rating range of the PDR and items with somewhat overlapping content (e.g., “Builds trust” and “Creates a positive environment”). A similar principal component analyses for the peer version of second-year students yielded the same single component structure (first two eigenvalues = 4.00 and 0.57), indicating that this structure was not merely a function of the first-year PDRs and instead a function of the measure. It was also noted that the Cronbach's alpha of first-year PDRs for instructor, peer, and self-ratings was very high (0.97, 0.96, and 0.96, respectively), which indicated the overall mean value of the PDR by rater was a reasonably consistent metric for use in further analyses. We proceeded

with analyses inputting the average across the 23 items by rater.

Primary Analyses

The R statistical package version 3.3.0, lavaan package (Rosseel, 2012) was used for all latent growth curve (LGC) modeling. An LGC model was fit to the instructor PDR, cadet PDR, self-rated PDR, academic GPA (academic performance score), military GPA, (military performance score or MPS), and physical GPA (physical performance score) by year. Intercept and linear slope terms estimated initial estimates and developmental changes (Figure 1). The first coefficient of the slope factor was fixed to 0 so that the intercept would capture the initial level, that is, the value during the first year. Race, (coded into three groups: White, Black, and other), gender, recruited National Collegiate Athletic Association athlete (binary), SAT, and the two honor violation categories, Lying and Cheating (binary) were entered as exogenous, time-invariant predictors to the latent slopes and intercepts.

This model provided an acceptable fit to the data, (CFI = 0.96, SRMR = 0.035, RMSEA = 0.074). Table 2 presents the latent slopes and intercepts, and the relationships of these parameters to the demographic and behavioral exogenous predictors, and Figure 2 shows the significant interrelationships between slopes and intercepts, as well as two exogenous categories race and gender (see Table 2 for the full model of estimates). A positive slope, indicating significant improvement, was found for instructor and self-rated PDR, and military and physical GPA. Statistically significant race effects (White > Black and other) were observed for the intercepts for cadet PDR, self-rated PDR, and all three GPAs. For the instructor PDRs, significant slope terms indicated that Black cadets are not rated as improving over time as much as other races are. For self PDR, male and White cadets rate themselves initially higher than Black or female cadets do; slope terms suggested that women

eventually match their male peers in self-perception, but Black cadets do not. Female cadets had higher academic and lower military GPA than male cadets; slopes indicated that women improved their military GPA faster than men did, but the gender difference in academic GPA was consistent across years. Finally, an honor conviction for lying was more highly associated with GPA intercepts, whereas a conviction for cheating was associated with negative GPA slopes.

Admission SAT score was included in the model and was significantly related to all latent constructs except cadet and self PDR slope, cadet PDR intercept, and physical GPA. As SAT is a moderate to strong correlate of academic success (e.g., Cohn et al., 2004; Fleming, 2002; Young, 2001), this finding suggested that an achievement factor is a major component of both success, as measured by GPA, and character and professional development, as measured by the PDR. However, it appears that this factor does not account for variation in PDRs completed by cadet peers. It is also important to note that, even when SAT is included in the model, White cadets have higher initial GPAs across all three pillars, receive higher PDR scores from their peers, over time are rated as improving more by their instructors, and score themselves higher on the PDR. This pattern may suggest implicit bias in these measures. Similarly, recruited athletes show several disparities across GPAs, and are rated lower on the PDR by their peers. Some of the GPA differences might be understood as the increased time commitment required of National Collegiate Athletic Association-level athletes, which may detract from time available for academics and military-relevant assignments and duties. Implicit and/or explicit bias against these students is another possibility.

Table 3 shows the coefficient matrix among the latent slopes and intercepts, which describes how the PDRs and GPAs relate to each other (i.e., the covariances between latent effects). All latent intercepts were positively related to each other, suggesting a general

TABLE 2
Coefficients for the Omnibus LGC for Model and by Covariate

			<i>Race- Black^{a, b}</i>	<i>Race- Other^{a, c}</i>	<i>Gender^d</i>	<i>Athlete^e</i>	<i>SAT</i>	<i>Honor- Cheat^f</i>	<i>Honor- Lying</i>
PDR									
Instructor	Intercept	2.81***	0.02	-0.01	-0.02	-0.05	0.11***	0.00	0.15
	Slope	0.28***	-0.06*	-0.03	-0.04*	0.02	-0.04***	-0.02	-0.17
Cadet	Intercept	3.22***	-0.12***	-0.05*	0.01	-0.09***	-0.00	0.01	-0.06
	Slope	0.02	0.03	-0.01	-0.01	0.03*	0.01	-0.05	-0.03
Self- rating	Intercept	3.07***	-0.12***	-0.05	0.09***	-0.11***	-0.04***	0.07	-0.07
	Slope	0.06*	0.03	-0.01	-0.04**	0.01	0.00	-0.02	0.03
GPA									
Academic	Intercept	3.30***	-0.13**	-0.10**	-0.07*	-0.15***	0.34***	-0.25**	-0.15
	Slope	0.02	-0.02	-0.00	-0.01	0.03***	-0.03***	-0.03	-0.06*
Military	Intercept	3.14***	-0.28***	-0.17***	0.08	-0.23***	0.13***	-0.16*	-0.13
	Slope	0.07**	0.00	0.00	-0.04***	0.04***	-0.04***	-0.02	-0.10*
Physical	Intercept	2.65***	-0.24***	-0.13***	0.05	0.14***	0.03	0.04	-0.20
	Slope	0.10***	0.02	-0.00	-0.02*	0.00	0.01**	-0.06**	-0.05

Notes: ^aFor LGC models, categorical variables are coded as n-groups - 1 binary indicators. ^bBlack versus White and Other. ^cOther versus White and Black. ^dPositive coefficient indicates Men > Women. ^ePositive coefficient indicates Recruited Athletes > Nonathletes. ^fPositive coefficient indicates cadet convicted of honor violation > cadet never convicted.

* $p < .05$. ** $p < .01$. *** $p < .001$.

aspect of initial “success” or “struggle” that cuts across program GPAs and PDRs. Instructor PDR slopes and GPA slopes were interrelated, whereas cadet-rated PDR slope was related to academic and military GPA slope only, and self-rated slope was not related to any other slope.

GPA slopes were negatively related to their intercept (e.g., high academic GPA was associated with lower slope). This finding suggests that cadets who struggle initially improve, whereas exceptional first-year cadets have only a small potential to increase their GPA. This finding was not the case for the PDR (except for the self-ratings), indicating more diversity across cadets in how initial score relates to character and professional growth.

Across latent variables, the military GPA was related to nearly every other latent variable (except military slope to self-rated PDR

slope, as noted above, and military intercept to physical GPA slope). This finding implies that performance and growth in military GPA is central to a cadet’s performance across domains. Among PDR raters, instructor-completed PDR slopes were inter-related to many other constructs, whereas self-rated PDR slope was related to a smaller set of latent constructs and was not related to any other PDR slopes.

DISCUSSION

The present study was designed to examine developmental trajectories in professional development and performance in cadets within the specific setting of the USMA, and to understand how these trajectories might be influenced by race, gender, and athlete status. We also evaluated how these patterns related

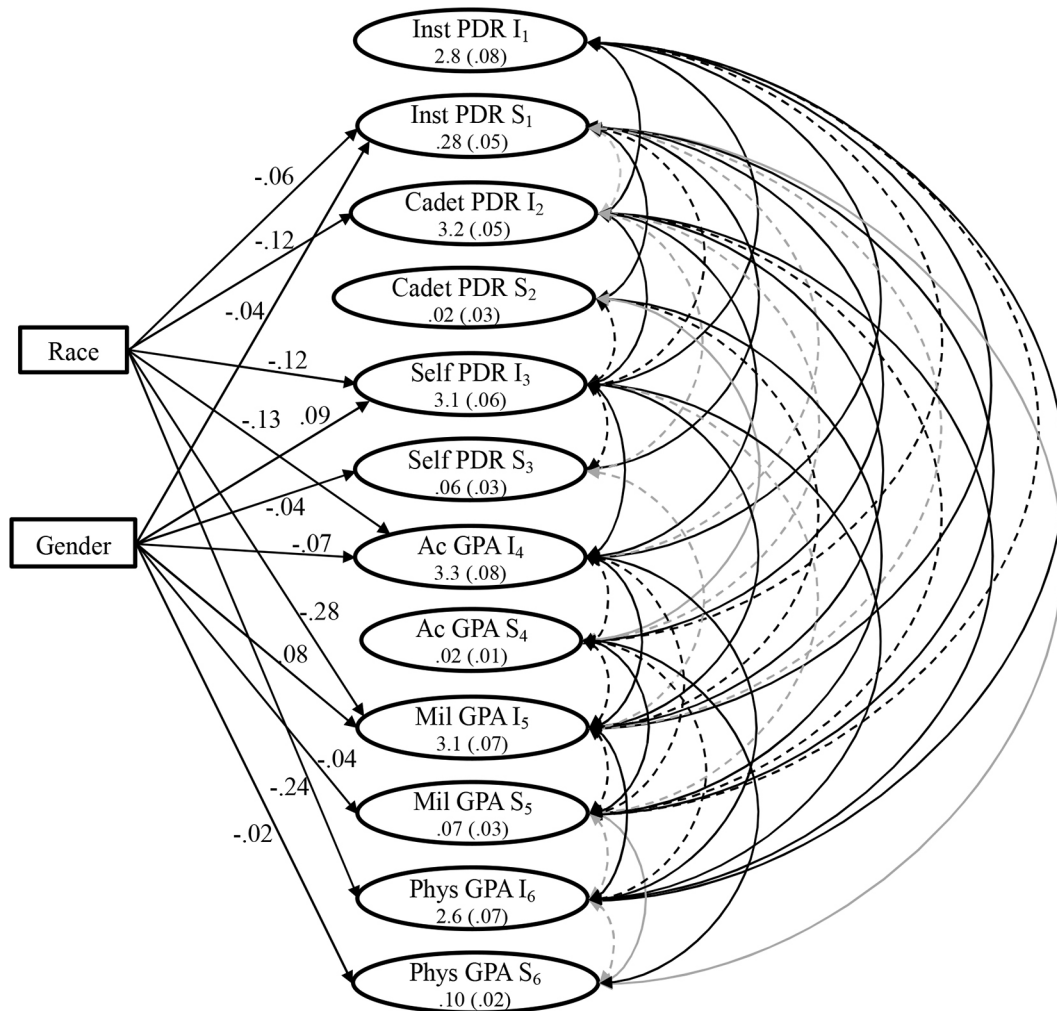


FIGURE 2

Abridged growth model results. Only race (Black versus White cadet comparison) and gender are shown for the time-invariant predictors for clarity, and unstandardized β estimates are given ($p < 0.05$). The latent estimates are shown with *SE* in parentheses. Positive β for race indicate White > Black, and positive β for gender indicates male > female. Given the number of slopes and intercepts, covariance estimates with $p < 0.001$ are shown with double-headed black arrows, and $0.001 < p < 0.05$ are shown with gray arrows. Dashed arrows indicate negative covariation, solid indicate positive covariation. See Tables 2 and 3 for the full list of model estimates.

to negative behavioral outcomes (i.e., honor violations). We were particularly interested in USMA's professional assessment instrument, the PDR, an analog of the professional evaluation system used in the larger Army. The PDR evaluates cadets on facets of character import-

ant to the development strategy at USMA. Understanding this instrument, its correlates and trajectories, and revealing how it relates to measures of performance such as grades allowed us to examine how professionalism develops and operates in this context. The

TABLE 3
Latent Slope-Intercept Covariances for the Omnibus LGC

	Slopes						Intercepts					
	PDR Score						GPA					
	Instructor	Cadet	Self	Academic	Military	Physical	Instructor	Cadet	Self	Academic	Military	Physical
Slopes												
PDR												
Instructor	—											
Cadet	.002***	—										
Self	.002***	.001	—									
GPA												
Academic	.006***	.001**	.001	—								
Military	.007***	.005***	.001	.003***	—							
Physical	.003**	.001	.001	.002***	.002**	—						
Intercepts												
PDR												
Instructor	-.007	-.001	-.001	-.006	-.01***	-.001	—					
Cadet	-.006*	.004	-.003**	-.002**	-.006***	-.001	.03***	—				
Self	-.01***	-.006***	-.02***	-.001	-.005**	-.002	.03***	.02***	—			
GPA												
Academic	-.01*	-.002	.001	-.01	-.02***	.001	.06***	.03***	.02***	—		
Military	.01**	-.007***	-.004*	-.005***	-.02***	-.002	.06***	.05***	.03***	.11***	—	
Physical	-.003	.001	-.002	-.004***	-.004*	-.004*	.03***	.03***	.02***	.07***	.06***	—

* $p < .05$. ** $p < .01$. *** $p < .001$.

specificity of our findings to the USMA context is certainly the case. Nevertheless, the fact that our findings represent an approach to assessing alignment between the professional values within an institution and the course of character development in the setting may have resonance (and perhaps specific applicability) to student-institution relations operating in other service academies, other educational settings with a character focus, or in other professional-training programs that value growth in both performance and character (e.g., law enforcement; Blumberg et al., 2016).

Preliminary analyses suggested that the 23 items of the PDR that describe the Army professional ethic (U.S. Department of the Army, 2012) cluster into a single underlying factor. The structure of the PDR measure is broad (many constructs assessed), but also flat (narrow range of response options), which is similar to other military evaluation instruments used for promotion and professional development (e.g., the Officer Evaluation Report). The results of the current study implied that those tools might also suffer from similar psychometric problems; indeed they have come under criticism (e.g., Chapman, 2006; Kiter, 1998) as both tools of promotion and advising (Fallesen et al., 2011). Future research on evaluation measures for military service members, as well as other programs and institutions with a character focus, should keep psychometric quality in mind when developing measures, so that the underlying constructs driving those scores can be investigated. It is also vital for investigations of service-member development or other military outcomes to consider the limitations of the evaluation tools that are implicated in the work.

Overall, the trajectories of professional values development and performance showed considerable evidence of growth: the negative relationship between intercept and slope indicated that cadets who have a difficult first year go on to improve considerably in years two and three. The performance and professional development metrics were highly related to

each other in terms of their slopes and intercepts; this result suggested that both are important to the overall development of cadets. Although the PDR's scale was intentionally developmental, with higher scores expected of older students, the program grades have no such expectation, and the covariance between performance and professional values indicated they grow together in meaningful ways across programs and observers (rating sources). This evidence of resilience, persistence, and growth is important insofar as these values are explicitly championed by USMA (e.g., USMA, 2018), and also because of the importance of resilience generally in a military setting (e.g., Tsai et al., 2016).

However, this evidence for positive developmental change in the face of challenge was coupled with evidence of systematic group differences that persisted despite inclusion of scores pertinent to academic achievement (i.e., SAT). These group differences existed across both performance and outcomes and rating scores. We found group differences in military GPA by race, athlete status, and gender, indicating that, even when SAT was accounted for, White, male, and nonathlete cadets received higher military program grades. The military program GPA integrates military-relevant classwork and subjective leadership ratings, an important metric at USMA and one that has strong connections to outcomes in the general Army (Bartone et al., 2002; Butler, 1976). Importantly, Black cadets had a lower initial peer PDR, self PDR, and lower GPA in all training programs, again, despite inclusion of SAT score, their slopes were not significantly different, which implied this effect remained throughout at least their first three years. In comparison, women have a lower initial military GPA but a higher slope, which suggested that they recovered from the initially lower start. Athletes show a similar pattern of lower initial military and academic GPA, but increased growth over time.

Full delineation of why these group differences exist was outside the scope of this analysis; however, a number of possibilities are

important to consider for future investigations. Evaluating the nature and source of differences is vital to ameliorating disparities in disadvantaged groups (Devine et al., 2017; Forscher & Devine, 2016). For the PDR, the complex interaction between rater and ratee (peer and student, or instructor and student) might result in group disparities due to implicit effects, self-fulfilling prophecies, confirmation bias, explicit beliefs, systemic bias, stereotype threat, and so on, each with a vast literature. The relatively inflated self-rating of men as compared to women is consistent with past work (e.g., Furnham, 2001) and suggests increased feelings of belonging or self-affirmation that tend to be lower in Black and female students in White, male majority collegiate environments (e.g., Hausmann et al., 2007). The racial achievement gap (i.e., lower GPA *despite* SAT score) is consistent with previous work, and has previously been shown to be associated with stereotype threat and other internalized beliefs about abilities (Walton & Spencer, 2009).

Overall, these influences may be highly context-dependent (Fleming, 2002), and underscored when the immediate context is stereotype-inconsistent (Lyness & Heilman, 2006). In the case of the PDR and the military program GPA, group differences might be particularly vulnerable to implicit forces because of the relatively imprecise evaluation criteria, due to multicollinearity in the former case and impression-based leadership evaluation in the latter (Fazio & Olson, 2014; Hansbrough et al., 2015). Possible sources of bias in existing professional and military evaluation systems warrants further empirical attention, as does interventions aimed at reducing these effects. Future work should evaluate how the implicit and explicit attitudes of raters and ratees might shape how and when group differences emerge, and how race or gender-based differences in feedback might later impact performance, so that more targeted interventions and precise evaluations can be implemented. (e.g., P. Williams et al., 2016).

These group differences are important, not only for institution-level understanding of the experiences and dynamics of cadets of color, but also because of the possible implications of these grades and evaluations on their future careers. Adverse experiences such as negative evaluations and earning lower grades than expected might compromise cadets' institutional trust in USMA and the Army generally, and research indicates such institutional trust may be fragile for Black students in particular (Yeager et al., 2017). Importantly, the relationship between performance, institutional trust, and engagement should be critically studied in diverse groups, if the Army is to continue their mission to recruit and retain a diverse force.

Overall, our results suggest that USMA is an environment of character development and resilience, but also systematic group disparities. Although the variation in professional development and performance metrics by race and gender have important implications for cadets' development, the current study is not without limitations. Importantly, we did not have the statistical power to allow race, gender, and athlete status to interact, or to investigate models by these demographic subgroups. We were also unable to investigate possible disparities among racial or ethnic groups with less representation at USMA, such as biracial, Latinx, or Asian cadets. However, we would hypothesize important subgroup differences that future investigations should consider, such as athletes of color, or women of color (e.g., Sellers, 1992).

Although we were able to investigate two primary indicators of development, success, and leadership at USMA—professional development and performance scores—there are many other metrics at USMA to consider in terms of trajectories and their interconnections. Notably, both the physical and military GPAs are a combination of classroom grades and performance scores in these domains. Investigating the components of these scores for their relevant power in determining cadet outcomes at USMA and in the larger Army will be important to translate trajectory find-

ings into specific recommendations. We analyzed the metrics as a whole for the current investigation in order to better understand the large-scale values of USMA in context, and also to see how metrics that the students most care about, notably program GPAs, interact with each other and grow with time. However, to fully investigate the processes of how various measures of professional growth, character, and performance are related, a more fine-grained analysis would be warranted. Overall, continuing to broaden the understanding of USMA's 47-month experience from the perspective of diverse groups of cadets is vital to characterize this developmental model of professionalism.

CONCLUSIONS

The question of what to assess when evaluating students and trainees is a key issue for universities and other training institutions, especially regarding character (Colby et al., 2003). Research on the links between character and professional growth has focused on the importance of moral and civic education, and in educating professionals for civic responsibility, stressing the roles of both the individual and the institution in professional development (D. A. Williams et al., 2005). However, relatively less is known about the relationship between performance mastery and professionally-relevant virtues, even in educational settings with clear character development missions, such as the U.S. services academies. This contextually derived character investigation can provide USMA, the larger military environment, and other institutions that explicitly champion character education with a window into the successes and liabilities of their curriculum and evaluation procedures. We hope that this introductory investigation of the primary components of USMA's professional preparation model will spark further evaluation of the nuances of this and similar systems.

Acknowledgments: We sincerely thank the Office of Institutional Research at USMA and the rest of the Project Arete team for their help and support. Preparation of this article was supported by a grant from the Templeton Religion Trust (to Kristina Schmid Callina and Richard M. Lerner).

NOTE

¹Demographic identifiers used were those collected by USMA, which use male or female to identify gender, and the following options for race/ethnicity: White/Caucasian, Black, Asian, Hispanic, American Indian, and other. Race and ethnicity are indicated together via this item. For the purpose of analyses, Asian, Hispanic, American Indian, and other were all coded as "Other."

REFERENCES

- Adler, P., & Adler, P. A. (1985). From idealism to pragmatic detachment: The academic performance of college athletes. *Sociology of Education*, 58(4), 241. <https://doi.org/10.2307/2112226>
- Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, 91(4), 652–661. Retrieved from <http://psycnet.apa.org/buy/2006-12810-006>
- Arbeit, M. R. (2016). "Make sure you're not getting yourself in trouble": Building sexual relationships and preventing sexual violence at the U.S. Military Academy at West Point. *The Journal of Sex Research*, 54(8), 949–961. <https://doi.org/10.1080/00224499.2016.1207055>
- Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice; Would Jesse Jackson "fail" the implicit association test? *Psychological Inquiry*, 15(4), 257–278. https://www.tandfonline.com/doi/pdf/10.1207/s15327965pli1504_01
- Baldwin, J. N. (1996). Female promotions in male-dominant organizations: The case of the United States Military. *The Journal of Politics*, 58(4), 1184–1197. <https://doi.org/10.2307/2960156>

- Bartone, P. T., Snook, S. A., & Tremble, T. R. (2002). Cognitive and personality predictors of leader performance in West Point Cadets. *Military Psychology, 14*(4), 321–338. https://doi.org/10.1207/S15327876MP1404_6
- Bass, B. M., & Yammarino, F. J. (1991). Congruence of self and others' leadership ratings of naval officers for understanding successful performance. *Applied Psychology, 40*(4), 437–454. <https://doi.org/10.1111/j.1464-0597.1991.tb01002.x>
- Berkowitz, M. W. (2012). Moral and character education. In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA handbooks in psychology. APA educational psychology handbook, Vol. 2. Individual differences and cultural and contextual factors* (pp. 247–264). American Psychological Association. <https://doi.org/10.1037/13274-010>
- Biernat, M. (2003). Toward a broader view of social stereotyping. *American Psychologist, 58*(12), 1019–1027. <https://doi.org/10.1037/0003-066X.58.12.1019>
- Blumberg, D. M., Giromini, L., & Jacobson, L. B. (2016). Impact of police academy training on recruits' integrity. *Police Quarterly, 19*(1), 63–86.
- Boe, O., Bang, H., & Nilsen, F. A. (2015). Experienced military officer's perception of important character strengths. *Procedia - Social and Behavioral Sciences, 190*, 339–345. <https://doi.org/10.1016/j.sbspro.2015.05.008>
- Boldry, J., Wood, W., & Kashy, D. A. (2001). Gender stereotypes and the evaluation of men and women in military training. *Journal of Social Issues, 57*(4), 689–705. https://dornsife.usc.edu/assets/sites/545/docs/Wendy_Wood_Research_Articles/Gender_Differences_in_Social_Behavior/Boldry.Wood.Kashy.2001_Sex_stereotypes_and_the_evaluation_of_men_and_women_in_military_training.pdf
- Bowman, W. R., & Mehay, S. L. (1999). Graduate education and employee performance: Evidence from military personnel. *Economics of Education Review, 18*(4), 453–463. [https://doi.org/10.1016/S0272-7757\(99\)00014-X](https://doi.org/10.1016/S0272-7757(99)00014-X)
- Boyce, L. A., & Herd, A. M. (2003). The relationship between gender role stereotypes and requisite military leadership characteristics. *Sex Roles, 49*(7/8), 365–378. <https://doi.org/10.1023/A:1025164221364>
- Burk, J., & Espinoza, E. (2012). Race relations within the US Military. *Annual Review of Sociology, 38*(1), 401–422. <https://doi.org/10.1146/annurev-soc-071811-145501>
- Butler, R. (1999). Why Black officers fail. *Parameters, 29*(3), 54.
- Butler, R. P. (1976). Relationships between college performance and success as an Army officer. *Journal of Vocational Behavior, 9*(3), 385–391. [https://doi.org/10.1016/0001-8791\(76\)90067-1](https://doi.org/10.1016/0001-8791(76)90067-1)
- Callina, K. S., Burkhard, B., Schaefer, H. S., Powers, J., Murray, E. D., Kobylski, G., Ryan, D.M., Kelly, D., Matthews, M.D., & Lerner, R. M. (2019). Character in context: Character structure among United States Military Academy cadets. *Journal of Moral Education, 48*(4), 439–464.
- Callina, K. S., Ryan, D., Murray, E. D., Colby, A., Damon, W., Matthews, M. D., & Lerner, R. M. (2017). Developing leaders of character at the United States Military Academy: A relational developmental systems analysis. *Journal of College and Character, 18*(1), 9–27.
- Carrell, S. E., Fullerton, R. L., & West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics, 27*(3), 439–464. <https://doi.org/10.1086/600143>
- Chapman, D. P. (2006). It's time for a new round OER reform. *Military Review, 86*(5), 100. <https://www.questia.com/library/journal/1G1-152935287/it-s-time-for-a-new-round-oer-reform>
- Chen, M., & Bargh, J. A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology, 33*(5), 541–560. <https://doi.org/10.1006/JESP.1997.1329>
- Cohen, T. R., & Morse, L. (2014). Moral character: What it is and what it does. *Research in Organizational Behavior, 34*, 43–61. <https://doi.org/10.1016/j.riob.2014.08.003>
- Cohn, E., Cohn, S., Balch, D. C., & Bradley, J. (2004). Determinants of undergraduate GPAs: SAT scores, high-school GPA and high-school rank. *Economics of Education Review, 23*(6), 577–586. <https://doi.org/10.1016/J.ECONEDUREV.2004.01.001>
- Colby, A., & Sullivan, W. M. (2008). Ethics teaching in undergraduate engineering education. *Journal of Engineering Education, 97*(3). <https://search.proquest.com/docview/217957230/full-textPDF/90A4FDA7AC7F4086PQ/?accountid=14678>

- Colby, A., Ehrlich, T., Beaumont, E., & Stephens, J. (2003). *Educating citizens: Preparing America's undergraduates for lives of moral and civic responsibility* (6th ed.). John Wiley & Sons. <https://doi.org/10.1080/00091380309604127>
- Crandall, C. S., & Eshleman, A. (2003). A justification–Suppression model of the expression and experience of prejudice definition of prejudice. *Psychological Bulletin*, 129(3), 414. <https://doi.org/10.1037/0033-2909.129.3.414>
- Currier, J. M., Holland, J. M., Drescher, K., & Foy, D. (2015). Initial psychometric evaluation of the Moral Injury Questionnaire–Military Version. *Clinical Psychology & Psychotherapy*, 22(1), 54–63. <https://doi.org/10.1002/cpp.1866>
- Cushen, D. B., Doty, J. P., & Toffler, P. A. (2012). *Developing trustworthy commissioned officers: Transcending the honor codes and concept*. <http://www.dtic.mil/dtic/tr/fulltext/u2/a573152.pdf>
- Darley, J. M., & Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1), 20–33. <https://doi.org/10.1037/0022-3514.44.1.20>
- Dennis, M., Francis, D. J., Cirino, P. T., Schachar, R., Barnes, M. A., & Fletcher, J. M. (2018). Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society*, 15, 331–343. <https://doi.org/10.1017/S1355617709090481>
- Devine, P. G., Forscher, P. S., Cox, W. T. L., Kaatz, A., Sheridan, J., & Carnes, M. (2017). A gender bias habit-breaking intervention led to increased hiring of female faculty in STEM departments. *Journal of Experimental Social Psychology*, 73, 211–215.
- Donaldson, S. I., & Grant-Vallone, E. J. (2002). Understanding self-report bias in organizational behavior research. *Journal of Business and Psychology*, 17(2), 245–260. <https://doi.org/10.1023/A:1019637632584>
- Dorans, N. J. (1999). Correspondences between ACT™ and SAT® I scores. *ETS Research Report Series*, 1999(1), i–18. <https://doi.org/10.1002/j.2333-8504.1999.tb01800.x>
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and inter-racial interaction. *Journal of Personality and Social Psychology*, 82(1), 62. <https://doi.org/10.1037/0022-3514.82.1.62>
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087–1101. <https://doi.org/10.1037/0022-3514.92.6.1087>
- Eitzen, D. S., & Purdy, D. A. (1986). The academic preparation and achievement of Black and White collegiate athletes. *Journal of Sport and Social Issues*, 10(1), 15–29. <https://doi.org/10.1177/019372358601000103>
- Eskreis-Winkler, L., Shulman, E. P., Beal, S. A., & Duckworth, A. L. (2014). The grit effect: Predicting retention in the military, the workplace, school and marriage. *Frontiers in Psychology*, 5, 36. <https://doi.org/10.3389/fpsyg.2014.00036>
- Fallesen, J. J., Keller-Glaze, H., & Curnow, C. K. (2011). A selective review of leadership studies in the US Army. *Military Psychology*, 23(5), 462.
- Fazio, R. H., & Olson, M. A. (2014). The MODE model: Attitude-behavior processes as a function of motivation and opportunity. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 155–171). Guilford Press.
- Fleming, J. (2002). Who will succeed in college? When the SAT predicts Black students' performance. *The Review of Higher Education*, 25(3), 281–296. <https://doi.org/10.1353/rhe.2002.0010>
- Forscher, P. S., & Devine, P. G. (2017). Knowledge-based interventions are more likely to reduce legal disparities than are implicit bias interventions. *Psychological Science Faculty Publications and Presentations*. <https://scholarworks.uark.edu/psycpub/5>
- Furnham, A. (2001). Self-estimates of intelligence: Culture and gender difference in self and other estimates of both general (g) and multiple intelligences. *Personality and Individual Differences*, 31(8), 1381–1405.
- Gaston-Gayles, J. L. (2004). Examining academic and athletic motivation among student athletes at a Division I university. *Journal of College Student Development*, 45(1), 75–83. <https://doi.org/10.1353/csd.2004.0005>
- Gralewski, J., & Karwowski, M. (2013). Polite girls and creative boys? Students' gender moderates accuracy of teachers' ratings of creativity. *The Journal of Creative Behavior*, 47(4), 290–304. <https://doi.org/10.1002/jocb.36>
- Greenhaus, J. H., Parasuraman, S., & Wormley, W. M. (1990). Effects of race on organizational experience, job performance evaluations, and

- career outcomes. *Academy of Management Journal*, 33(1), 64–86. <https://doi.org/10.2307/256352>
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17. <http://psycnet.apa.org/buy/2009-08950-006>
- Grunspan, D. Z., Eddy, S. L., Brownell, S. E., Wiggins, B. L., Crowe, A. J., & Goodreau, S. M. (2016). Males under-estimate academic performance of their female peers in undergraduate biology classrooms. *PLOS ONE*, 11(2), e0148405. <https://doi.org/10.1371/journal.pone.0148405>
- Hall, M. P., & Raimi, K. T. (2018). Is belief superiority justified by superior knowledge? *Journal of Experimental Social Psychology*, 76, 290–306. <https://doi.org/10.1016/J.JESP.2018.03.001>
- Hansbrough, T. K., Lord, R. G., & Schyns, B. (2015). Reconsidering the accuracy of follower leadership ratings. *The Leadership Quarterly*, 26(2), 220–237.
- Hanser, L. M., & Oguz, M. (2015). *United States Service Academy Admissions: Selecting for success at the Military Academy/West Point and as an officer*. <http://www.dtic.mil/docs/citations/ADA616062>
- Harris, G. L. A. (2009). Recruiting, retention, and race in the military. *International Journal of Public Administration*, 32(10), 803–828. <https://doi.org/10.1080/01900690902905485>
- Hausmann, L. R., Schofield, J. W., & Woods, R. L. (2007). Sense of belonging as a predictor of intentions to persist among African American and White first-year college students. *Research in Higher Education*, 48(7), 803–839.
- Hershberg, R. M., Rubin, R. O., Johnson, S. K., Callina, K. S., & Lerner, R. M. (2016). Perceptions of character development at a trade college: Triangulating student, alumni, administrator, and teacher perspectives. *Research in Human Development*, 13(2), 157–173. <https://doi.org/10.1080/15427609.2016.1164553>
- Hopkins, R., & Williams, W. A. (2013). Inequities in the officer promotion system. In I. Hampton (Ed.), *The Black Officer Corps: A history of Black military advancement from integration through Vietnam* (pp. 62–68). Routledge.
- Jeynes, W. H. (2017). A meta-analysis on the relationship between character education and student achievement and behavioral outcomes. *Education and Urban Society*, 51(1), 33–71. <https://doi.org/10.1177/0013124517747681>
- Kane, T. J. (2011). The entrepreneurial Army: A survey of West Point graduates. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1734594>
- Kelly, D. R., Matthews, M. D., & Bartone, P. T. (2014). Grit and hardiness as predictors among West Point cadets. *Military Psychology*, 26(4), 327–342. <http://www.hardiness-resilience.com/docs/Battlebook-129-pages-2AUG2015.pdf#page=104>
- Kitter, D. P. (1998). *The U.S. Army Officer Evaluation Report: Why are we writing to someone who isn't reading?* <http://www.dtic.mil/docs/citations/ADA398598>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin*, 129(4), 522–544. <https://doi.org/10.1037/0033-2909.129.4.522>
- Lai, C. K., Hoffman, K. M., & Nosek, B. A. (2013). Reducing implicit prejudice. *Social and Personality Psychology Compass*, 7(5), 315–330. <https://doi.org/10.1111/spc3.12023>
- Le, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, 25(1), 1–18. 10.18637/jss.v025.i01
- Lerner, R. M. (2018). Character development among youth: Linking lives in time and place. *International Journal of Behavioral Development*, 42(2), 267–277.
- Lerner, R. M., & Callina, K.S. (2014). The study of character development: Towards tests of a relational developmental systems model. *Human Development*. <https://doi.org/10.1159/000368784>
- Lyness, K. S., & Heilman, M. E. (2006). When fit is fundamental: Performance evaluations and promotions of upper-level female and male managers. *Journal of Applied Psychology*, 91(4), 777–785. <https://doi.org/10.1037/0021-9010.91.4.777>

- Maddi, S. R., Matthews, M. D., Kelly, D. R., Villarreal, B. J., Gundersen, K. K., & Savino, S. C. (2017). The continuing role of hardiness and grit on performance and retention in West Point cadets. *APA*. <https://doi.org/10.1037/MIL0000145>
- Marini, J. P., Shaw, E. J., & Young, L. (2016). Using old and new SAT® scores for admission: A closer look at concordant scores in predictive models (Research Report 2016-17). *College Board*. <https://files.eric.ed.gov/fulltext/ED572569.pdf>
- Matthews, M. D., Ender, M. G., Laurence, J. H., & Rohall, D. E. (2009). Role of group affiliation and gender on attitudes toward women in the military. *Military Psychology*, 21, 241–251. <https://doi.org/10.1080/08995600902768750>
- McCaslin, M. (2016). Commentary: The co-regulation of cultural, social, and personal sources of influence on student emotional adaptation and achievement. In K. R. Wentzel & G. B. Ramani (Eds.), *Handbook of social influences in school contexts: Social-emotional, motivation, and cognitive outcomes* (pp. 329–346). Taylor & Francis.
- Moore, D. W., & Trout, B. T. (1978). Military advancement: The visibility theory of promotion. *American Political Science Review*, 72(2), 452–468. <https://doi.org/10.2307/1954104>
- Mueller, U., & Mazur, A. (1996). Facial dominance of West Point cadets as a predictor of later military rank. *Social Forces*, 74(3), 823–850. <https://doi.org/10.1093/sf/74.3.823>
- Pazy, A., & Oron, I. (2001). Sex proportion and performance evaluation among high-ranking military officers. *Journal of Organizational Behavior*. Wiley. <https://doi.org/10.2307/3649615>
- Peterson, C., & Seligman, M. E. (2004). *Character strengths and virtues: A handbook and classification*. Oxford University Press.
- Pigott, R. L., & Cowen, E. L. (2000). Teacher race, child race, racial congruence, and teacher ratings of children's school adjustment. *Journal of School Psychology*, 38(2), 177–195. [https://doi.org/10.1016/S0022-4405\(99\)00041-2](https://doi.org/10.1016/S0022-4405(99)00041-2)
- Prividera, L. C., & Howard, J. W. I. (2006). Masculinity, Whiteness, and the warrior hero: Perpetuating the strategic rhetoric of U.S. Nationalism and the marginalization of women. *Women and Language*, 29(2), 29–37. <https://search.proquest.com/docview/198816262?pq-origsite=gscholar>
- Riegle-Crumb, C., & Humphries, M. (2012). Exploring bias in math teachers' perceptions of students' ability by gender and race/ethnicity. *Gender & Society*, 26(2), 290–322. <https://doi.org/10.1177/0891243211434614>
- Robins, R. W., & Pals, J. L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self and Identity*, 1(4), 313–336. Retrieved from <https://www.tandfonline.com/doi/pdf/10.1080/15298860290106805>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <http://www.jstatsoft.org/v48/i02/>
- Samuels, S. M., & Casebeer, W. D. (2005). A social psychological view of morality: why knowledge of situational influences on behaviour can improve character development practices. *Journal of Moral Education*, 34(1), 73–87. <https://doi.org/10.1080/03057240500049349>
- Sedlacek, W. E., & Adams-Gaston, J. (1992). Predicting the academic success of student-athletes using SAT and noncognitive variables. *Journal of Counseling & Development*, 70(6), 724–727. <https://doi.org/10.1002/j.1556-6676.1992.tb02155.x>
- Sellers, R. M. (1992). Racial differences in the predictors for academic achievement of student-athletes in Division I revenue producing sports. *Sociology of Sport Journal*, 9(1), 48–59. <https://doi.org/10.1123/ssj.9.1.48>
- Street, A. E., Stafford, J., Mahan, C. M., & Hendricks, A. (2008). Sexual harassment and assault experienced by reservists during military service: Prevalence and health correlates. *Journal of Rehabilitation Research and Development*, 45(3), 409–420. <https://doi.org/10.1682/JRRD.2007.06.0088>
- Tenenbaum, H. R., & Ruck, M. D. (2007). Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of Educational Psychology*, 99(2), 253–273. <https://doi.org/10.1037/0022-0663.99.2.253>
- Tsai, J., Mota, N. P., Southwick, S. M., & Pietrzak, R. H. (2016). What doesn't kill you makes you stronger: A national study of US military veterans. *Journal of Affective Disorders*, 189, 269–271.
- Turchik, J. A., & Wilson, S. M. (2010). Sexual assault in the U.S. military: A review of the literature and recommendations for the future.

- Aggression and Violent Behavior*, 15(4), 267–277. <https://doi.org/10.1016/J.AVB.2010.01.005>
- U.S. Department of the Army. (2012). *ADRP 6-22 Army Leadership*.
- United States Military Academy. (2016a). *Military program (Greenbook)*.
- United States Military Academy. (2016b). *Physical program whitebook*.
- United States Military Academy. (2018). *Developing leaders of character*.
- Van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *The Australian Journal of Advanced Nursing*, 25(4). Retrieved from <https://search.informit.com.au/documentSummary;dn=210155003844269;res=IELAPA>
- Van Iddekinge, C. H., Roth, P. L., Raymark, P. H., & Odle-Dusseau, H. N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, 97(3), 499–530. <http://psycnet.apa.org/doiLanding?doi=10.1037/a0021196>
- Von Culin, K. R., Tsukayama, E., & Duckworth, A. L. (2014). Unpacking grit: Motivational correlates of perseverance and passion for long-term goals. *The Journal of Positive Psychology*. <https://doi.org/10.1080/17439760.2014.898320>
- Walker, B. H., Sinclair, H. C., & MacArthur, J. (2015). Social norms versus social motives: The effects of social influence and motivation to control prejudiced reactions on the expression of prejudice. *Social Influence*, 10(1), 55–67. <https://doi.org/10.1080/15534510.2014.904247>
- Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. *Psychological Science*, 20(9), 1132–1139.
- Wells, R., Wolniak, G., Enberg, M., & Manly, C. A. (2016). Socioeconomic disparities in the use of college admission-enhancing strategies among high school seniors from the 1990s to 2000s. *Teachers College Record*, 118(9), 1–36.
- Williams, D. A., Berger, J. B., & McClendon, S. A. (2005). *Toward a model of inclusive excellence and change in postsecondary institutions*. Association of American Colleges and Universities. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.129.2597&rep=rep1&type=pdf>
- Williams, J. L., & Deutsch, N. L. (2016). Beyond between-group differences: Considering race, ethnicity, and culture in research on positive youth development programs. *Applied Developmental Science*, 20(3), 203–213. <https://doi.org/10.1080/10888691.2015.1113880>
- Williams, P., Kern, M. L., & Waters, L. (2016). Exploring selective exposure and confirmation bias as processes underlying employee work happiness: an intervention study. *Frontiers in Psychology*, 7, 878. <https://doi.org/10.3389/fpsyg.2016.00878>
- Yeager, D. S., Purdie-Vaughns, V., Hooper, S. Y., & Cohen, G. L. (2017). Loss of institutional trust among racial and ethnic minority adolescents: A consequence of procedural injustice and a cause of life-span outcomes. *Child Development*, 88(2), 658–676.
- Young, J. W. (2001). Differential validity, differential prediction, and college admission testing: A comprehensive review and analysis. Research Report No. 2001-6. *College Entrance Examination Board*. <https://eric.ed.gov/?id=ED562661>
- Yu, Y.-M. (2018). Comparing moral education models at a military academy in Taiwan. *Journal of Academic Ethics*, 1–21. <https://doi.org/10.1007/s10805-018-9301-x>