People show different patterns of behavior depending on whether they succeed or fail pursuing their long-term goals. Whereas some people are steadfast in their pursuit of such goals, sometimes over years or decades, other people abandon pursuit in the face of significant challenges to the attainment of such goals (Arco-Tirado et al., 2018). Recent research findings have spurred the evidence of the so-called non-cognitive abilities as predictors of individual performance and productivity in different life domains (Duckworth & Kautz, 2012; Heckman et al., 2006; Méndez & Zamarro, 2015). A pleiad of terms like qualities, traits, non-cognitive competencies, soft skills, or life skills are interchangeably used as predictors of psychological, social, economic, and physical wellbeing (Almlund et al., 2011; Borghans et al., 2008; Heckman & Kautz, 2012; Heckman et al., 2006; Méndez & Zamarro, 2015). The variability of the terms mentioned above can be the result of the lack of a shared definition of the “non-cognitive skills” concept, including measurement problems (Duckworth & Yeager, 2015; Fernández-Martín et al., 2018; García, 2014). In this view, the most common approach among different authors refer to this concept as a pattern of thought, feelings, and behaviors (Borghans et al., 2008). With this term we also refer to diverse attitudes, personality characteristics, and behaviors including self-perceptions (i.e., self-concept and self-efficacy), motivation, self-control, metacognitive...
strategies, social competencies (i.e., social and leadership skills), resilience, creativity, and perseverance (i.e., engagement and grit) (García, 2014; Gutman & Schoon, 2013).

Within the Big Five taxonomy of personality traits, grit shares content with the conscientiousness domain (Rimfeld et al., 2016) as reflected in its conceptual relatedness to several of the conscientiousness facets – orderliness, dependability, self-control, and industriousness (Duckworth & Eksreis-Winkler, 2015). Indeed, grit is viewed by some personality theorists as falling with the broad family of conscientiousness constructs (Roberts et al., 2014). Yet, despite its conceptual similarity to facets of conscientiousness, none of the conscientiousness facets fully captures the combination of passion and perseverance that characterizes grit (Arco-Tirado et al., 2018). In particular, the sustained interest in important long-term goals, a core feature of grit, is not evident in conscientiousness or its facets (Sartori et al., 2017). Thus, although the conceptual overlap between the broad conscientiousness domain and the relative narrow grit trait is substantial – conceptually speaking, grit includes a specific and unique focus on the pursuit of long-term, higher-order goals (Arco-Tirado et al., 2018).

Grit was initially measured with the Original Grit Scale (Grit-O) developed by Duckworth et al. (2007) aiming to capture the characteristic attitudes and behaviours of high-performing individuals, albeit without associating them with a specific domain in life. The confirmatory factor analysis showed a structural model with two first-order factors – each made up of six items – although with fit indices with room for improvement. Thus, two years later Duckworth and Quinn (2009) developed and validated a more efficient measurement of the grit construct, the Short Grit Scale (Grit-S). The resulting Grit-S scale (8 items) was built with a larger sample and showed better psychometric properties than the Grit-O scale (12 items), revealing fairly acceptable fit indexes (i.e., RMSEA = .076 and CFI = .96), an internal consistency of .82, .77, and .70, respectively, a test-retest reliability (12 months) stability over time, r = .68, and evidence of the predictive validity of the Grit-S scale, including the differential for each factor, on psychological and educational variables such as educational level and a lower number of degree changes in adults (see Duckworth and Quinn, 2009 for more details). Additionally, the Grit-S scale (Duckworth & Quinn, 2009) has also been adapted and validated to other contexts, like Turkish (Akn et al., 2011) and Japanese (Nishikawa et al., 2015).

The predictive validity of grit is evident from a number of studies focused on this non-cognitive ability yielding prospective evidences that higher scores on grit predict multiple variables, such as: (a) grade point average and hours spent watching television – inversely –, without controlling for other variables, and educational attainment and fewer career changes in adults and retention for West Point cadets, controlling for conscientiousness and other Big Five inventory factors (Duckworth & Quinn, 2009); (b) higher retention rates among military cadets after the first summer of training (Duckworth & Quinn, 2009); (c) higher teacher retention and effectiveness in low-income school districts (Eksreis-Winkler, Shulman, & Duckworth, 2014; Robertson-Kraft & Duckworth, 2014); (d) higher labor stability and higher professional efficacy (Duckworth et al., 2009; Eksreis-Winkler, Shulman, Beal et al., 2014); (e) longer marriage – only for men – (Eksreis-Winkler, Shulman, Beal et al., 2014); and (f) higher life expectancy (Kim & Lee, 2015). Therefore, the widespread enthusiasm for grit in the last decade is not strange, as a potentially novel predictor and determinant of performance.

Indeed, Duckworth et al. (2007) noted that grit accounted for an average of 4% of the variance in success outcomes, which corresponds to a correlation of $r = .20$, or what is called a “small-to-medium” effect. Other authors like Sheldon et al. (2015) compared grit to nine other positive personality characteristics in a year-long study of goal attainment and found that grit was the most reliable predictor of achieving personally meaningful goals over that period. These findings suggest that grit is an important determinant of success, although not the only one. In another publication by Duckworth et al. (2015) these authors identify many other factors, including opportunity, talent, luck, as well as other personality characteristics like curiosity, that also influence achievement outcomes. Additionally, if we put in perspective the magnitude of the grit effect found, as other authors like Roberts et al. (2007) point, we realize the importance and capacity of grit as a single characteristic to predict objective outcomes, since those results do not differ from those found related to other personality traits.

From the intervention and educational perspective, the analysis of grit as an outcome turns out to be as necessary and important as grit as a predictor. The scarcity of prediction research on grit is clearly noticeable in the literature and it is probably due to the fact that researchers are more interested in studying the predictive performance of different measures rather than the reverse approach. Thus, although grit as an outcome appears linked to educational (e.g., Egalite et al., 2016; Lyon, 2014), socio-demographics (e.g., Johnson, 2014), professional (e.g., Clark, 2016), or personal factors (e.g., Incantalupo-Kuhner, 2015) in the literature, no previous studies have analyzed systematically the quality of the evidences supporting those prospective results affecting our current knowledge on how grit develops.

Based on the fact that studies on grit have experienced a geometric growth in the last ten years, and the number of authors claiming for the role of grit as both predictor and outcome of further personal, professional, and/or educational development, it was necessary to conduct a systematic review and/or meta-analysis on grit based on international high-quality standards and protocols, because, systematic reviews are indispensable components in the chain of scientific information and key tools for evidence-based (Ioannidis, 2016). Therefore, the objective of this systematic review is to synthesize the results available on grit as predictor of educational, professional, and personal success as outcomes, as well as grit as outcome of educational, professional, and personal factors as predictors, following international high-quality standards for systematic reviews (i.e., The Campbell Collaboration, 2015), and the research questions to answer are: (1) what are the salient features of grit research?, (2) what are the evidences accounting for grit as predictor in different domains of success?, (3) what are the evidences accounting for grit as outcome?, and (4) what direction should research on grit take to strengthening future causal research?

**Method**

**Inclusion and Exclusion Criteria**

The review team developed a systematic review protocol following Campbell Systematic Reviews: Policies and Guidelines (The Campbell Collaboration, 2015). The access to this protocol is available upon request to the review team.

The operational characteristics of independent (predictor) and dependent (outcome) variable(s) were established as follows.

**Grit** (either as a predictor or outcome): “perseverance and passion for long-term goals” (Duckworth, 2006, p. 71). The measurement of grit is carried out via: (a) the Short Grit Scale (Grit-S) (Duckworth & Quinn, 2009), and (b) any of the translated versions of this versions in other languages, for example, Japanese (Nishikawa et al., 2015), Filipino (Datu et al., 2016), and Turkish (Akn et al., 2011). Grit-S scale was selected because of the improved psychometric properties compared to previous versions (i.e., Grit-O) (Duckworth & Quinn, 2009) and therefore to increase results comparability.

**Educational success** (either as a predictor or outcome): “accomplishment of widely valued goals” (Duckworth et al., 2007, p. 1087) as school/academic performance (i.e., dropout – as the reverse
of retention –, retention, grade point average – GPA –, achievement test scores, repetition – reversed –, and graduation) and satisfaction, measured through an standardized and/or objective quantitative procedure, usually a questionnaire or a structured interview.

**Professional success** (either as a predictor or outcome): “accomplishment of widely valued goals” (Duckworth et al., 2007, p. 1087) as performance (e.g., assigned a ranking to each teacher based on the academic gains of their students over the course of the academic year) (Duckworth et al., 2009), retention (Eskreis-Winkler, Shulman, Beal et al., 2014), and satisfaction, measured through standardized or objective quantitative procedure, usually a questionnaire or a structured interview.

**Personal success** (either as a predictor or outcome): “accomplishment of widely valued goals” (Duckworth et al., 2007, p. 1087) as psychological well-being (e.g., optimal psychological functioning and experience) and satisfaction (e.g., how happy they feel with their life based on a person's own criteria of what makes them satisfied) (Tiitinen, 2014), measured through standardized and/or objective quantitative procedure, usually a questionnaire or a structured interview.

There was not a particular eligible participant population for this research study.

The eligible research designs for this research study were: pre-experimental, experimental, quasi-experimental, and correlational and ex post facto, following the classification of Campbell and Stanley (1963).

The time period covered in this research study extends from 2006, when Duckworth defended her dissertation on grit, onwards until 2017.

No geographical and/or cultural restrictions were included except for the language of publication, which were English and Spanish.

**Search Strategies**

The literature search was based on a variety of sources in order to ensure that published and unpublished studies (“grey literature”) relevant to the review questions were included in the search process. The search process included a primary search, searching of databases, and a complementary search, searching other resources and unstructured searching of relevant websites, literature snowballing, and contacting experts. The review team used RefWorks to manage and document this process. This search was conducted during September 2017.

The primary search of databases included a wide range of platform and electronic databases: Proquest, OvidSP, EBSCOHOST, Web of Science, Scopus, The online platform for Taylor & Francis, and specialized databases (i.e., Research for Development, British Library for Development Studies, Science Direct, Springer Link, PsyCh, Wiley Online Library, Center for Longitudinal Studies, REDINED, REDUC, ISOC, Redalyc, and Dialnet).

The primary search was supplemented by a complementary search to identify further studies for inclusion: (a) reference lists of included studies and reference lists of relevant reviews were searched, (b) a general web search was conducted using Google Scholar to identify potential unpublished studies, (c) personal contacts with national and international researchers were made to identify unpublished reports and ongoing studies, (d) conference proceedings (i.e., Ovid MedlineR In-Process & Other Non-Indexed Citations, Conference Papers Index, Scopus, Web of Science, and The Conference Collection), (e) dissertations and theses databases (i.e., Networked Digital Library of Theses and Dissertations, ProQuest Dissertations and Theses Global, Electronic Theses Online Service, The Australian Digital Theses, and TESEO), (f) open access (i.e., OpenGrey EAGLE, GreyNet International-Grey Literature Network Service, National Technical Information Service, OpenSIGLE, Directory of Open Access Repositories, Ovid HealthSTAR Database, COPAC, and Urbadisc), (g) ongoing research (i.e., CORDIS, Regard database of Economic and Social Research Council, Centre for Review and Dissemination, NBER Working Papers, The Campbell Collaboration, and RePeC), and (h) relevant institutions and networks (i.e., American Institutes for Research, What Works Clearinghouse, EPPI Centre, Educational Evidence Portal, IZA World of Labor, Social Science Research Network, The Duckworth Lab, The Grit Institute, Strive Together, Edutopia, Character Lab, and GRITLab).

The search terms (in English and Spanish) reflected the inclusion criteria and tried to strike a balance between sensitivity and specificity. To ensure inclusion of papers, which do not specifically report their research design, population or geographical/cultural restrictions, the search excluded methodology, population, and geographical or cultural restrictions terms. We classified search terms according to two categories, which were combined using the Boolean operator “AND” to identify potentially relevant studies in each database (in all fields from 2006 onwards): (“grit” AND “long term goals” AND “long-term goals” AND “perseverance”) AND (“success” OR “performance” OR “academic” OR “school” OR “grade point average” OR “GPA” OR “achievement” OR “learn” OR “outcome” OR “retention” OR “effort” OR “education” OR “student” OR “dropout” OR “drop-out” OR “employ” OR “job” OR “job loss” OR “work” OR “unemployment” OR “under-employment” OR “under-employ” OR “under-employer” OR “satisfaction” OR “well being” OR “well-being” OR “happiness” OR “optimism” OR “triumph” OR “self”).

In the selection process we implemented the following tasks: (a) the first screening level was aimed at identifying and removing duplicate records and/or published before 2006, (b) the second screening level was aimed at identifying and removing those studies in which the grit term was used with a different meaning from Duckworth's grit concept, after further examination of the title and abstract, and (c) the third screening level was aimed at reading the full text versions of the studies to ascertain eligibility based on the remaining inclusion criteria, and exclusion criteria like the language of publication, type of publication (e.g., newspapers), studies not based on data gathered through Grit-S (Duckworth & Quinn, 2009), and/or studies that do not provide overall grit score (i.e., performing statistical analyses based on Grit sub-scales scores separately). A screening guide was used in the second and third screening level. Based on the review’s inclusion and exclusion criteria, discrepancies were resolved by further review of the respective titles, abstracts and full text, and discussion. The overall search and screening process were documented in a flow diagram (Figure 1).

Data and information were coded and extracted from included studies. For the analysis, the approach adopted was a narrative content analysis (Dochy, 2006), and eligible studies were coded on variables related to: (a) study methods (i.e., sampling technique and procedure, response rate/attrition, representativeness, research design, statistical analysis, and bias), (b) predictor variables (e.g., grit, age, educational level, performance, self-control, etc.), (c) outcome variables (e.g., grit, GPA, performance, cognitive skills, etc.), (d) characteristics of the subject samples of analysis (i.e., sample size, mean and range age, and gender), and (e) contextual features (i.e., author/s, country, publication year, and title).

**Results**

The initial search identified a total of 5,855 publications (3,901 in primary search, and 1,954 in a complementary search). After checking for duplicated publications and/or published before 2006 (first screening level), 2,248 publications were excluded. Then, the title and abstract of 3,607 publications were examined (second screening level) and 3,292 publications were excluded. Finally, after reading the full text versions of 315 publications (third screening level) a total of 78 publications (containing 90 studies) qualified for the analysis,
with 8 studies that were not obtained in full text despite repeated attempts to locate them. Out of those 78 publications, 3 included 2 studies, 3 included 3 studies, and 1 included 4 studies, so the final N was 90 studies. Figure 1 illustrates the flowchart for the literature search and screening.

![Flow Chart for the Literature Search and Screening.](image)

**Figure 1.** Flow Chart for the Literature Search and Screening.

### Features of Grit Research

First, we present the main features of the grit research to offer a clearer picture of the field, in order to answer our first research question.

**Contextual features.** The total of studies selected belong to 11 countries, 78 from the United States of America, 2 from South Africa, 2 from Canada, and 1 each from Turkey, South Korea, Norway, United Kingdom, Filipine, China, Holland, and Denmark. The studies were published between 2009 (n = 5) and 2017 (n = 3) (i.e., 2011 = 2; 2012 = 6; 2013 = 6; 2014 = 37; 2015 = 22; 2016 = 9). Of the reviewed studies, 41 were parts of journal publications, 29 were reported in dissertation theses, 8 in master’s theses, and 12 came from other sources such as conference proceeding (n = 6), working papers (n = 1), poster (n = 1) or institutional research reports (n = 4).

### Sample

The studies reviewed include a total 70,963 participants, with a sample size ranging from 15 participants to 15,874 participants. Gender distribution varies among studies, with 72 studies including samples composed by males and females, 5 studies with only males, and 13 studies failing to report this data. Studies samples are composed generally by several ethnic groups except for 3 studies composed only by African American high school, college, or university students, 1 by White-Hispanic university students, and 1 by Latino students. In relation to studies’ population the review yielded the following results: 3 military, 4 children (i.e., finalists in the 2006 Scripps national spelling bee and children with reading disorder, childhood ADHD), 13 elementary, middle or high school students, 1 adult student who returned to a specific charter high school to complete a high school diploma, 1 promise academy students, 35 undergraduate and postgraduate students, 1 doctoral student, 13 workers (e.g., teachers, principals, police detectives, mental health workers, employed students, etc.), 12 diverse populations (e.g., general population, adolescents, adults, adults aged 18–72 years, runners, etc.), and 7 mixed (e.g., undergraduate students, parents and grandparents, students and teachers).

### Design

In relation to the studies research design our review identified a total of 19 studies (21.1%) failing to explicitly mention this information, although it was eventually inferred by reviewers from the rest of the information reported. Thus, the studies methodological features regarding research design show the following distribution: (a) 6 experimental designs (6.6%), (b) 6 pre-experimental designs (6.6%), with 1 study including a static-group design, and 5 including one-group pretest-posttest design, (c) 2 ex post facto designs or retrospective studies (2.2%), and (d) 76 correlational designs or predictive and explanatory studies (84.6%), with 68 cross-sectional and 8 longitudinal.

All studies gathered data through the English original version of the Grit-S scale (Duckworth & Quinn, 2009), except for 1 study using the Turkish version of the Grit-S scale (Akın et al., 2011).

The dominant statistical analyses were bivariate (e.g., t-test, U-test, chi-square test, one-way analysis of variance) and multivariate (e.g., two-way analysis of variance, regression and path analysis), considering that studies using only descriptive or correlational analysis (i.e., Pearson, Spearman, and Kendall tau) were selected out from our analysis. Additionally, sampling techniques were distributed as follows: 84 non-probabilistic (e.g., purposive sampling, convenience sampling, snowball sampling) and 6 probabilistic (e.g., random sampling).

Finally, reporting bias were included in 78 studies (86.67%) but not in 12 studies (13.33%).

### Evidences of Grit as Predictor

In the case of grit as predictor our review found a total of 57 studies (see Table 1).

In relation to educational success a total of 25 studies were found. Grit predicts educational attainment among general population, university, middle and high school students. Additionally, grit predicts GPA among middle school students, high school students, undergraduate students, and non-traditional doctoral students. Furthermore, grit predicts retention among freshmen and candidates at the United States Military Academy, the National Guard Youth Challenge Program Cadets at the Washington Youth Academy, and the 24-day Army Special Operations Forces selection course, and potential dropping-out among foreign university students and career changes in general population.

In relation to grit as predictor of professional success a total of 11 studies have been identified. Grit predicts objective career success like career status and salary, retention in the workplace, or teaching effectiveness, but not subjective career success, like career or job satisfaction, task performance or turnover intention. Additionally, grit predicts well-being, burnout, or stress in the workplace, optimism of principals, and work engagement among police detectives.

Finally, another set of 26 studies focused on personal success variables. For analyzing and synthesizing purposes, the different outcome variables were grouped into five categories (i.e., cognitive skills, engagement, mental health, physical health, and relationships) with two of them accumulating most of the studies and evidences. That is the case of cognitive skills (n = 11), representing quite diverse outcomes related to performance like spelling, literacy, expectations, and problem solving (e.g., anagrams, tasks that are based on clicking the computer mouse or solving math problems), and engagement as beliefs about intelligence. Furthermore, grit predicts subjective reports of psychological health care management skills among
university students, satisfaction with life in general populations, and frequency and intensity of spiritual and transcendent experiences happening within the context of everyday life. Grit also has indirect effect on suicide attempts, but direct effect of grit on suicide attempts was not significant. Grit is positively related to subjective reports of physical health, as well as exercise behavior, with moderate and high intensity, but neither variable was related to low intensity exercise, among university students, staff, and faculty. Finally, grit predicts romantic relationship satisfaction among adults, and grit-gender interactions explain variance in the retention marital status.

**Evidences of Grit as Outcome.** In the case of grit as outcome our review found a total of 56 studies (see Table 2). Out of those, 28 studies target socio-demographic variables as predictors of grit. Results show that age and gender/sex predict grit level among students (e.g., university, middle, and high school) and general population, with older students showing higher level of grit and women inclined to report higher levels of grit. Concurrently, age and gender/sex are rejected by some of these studies as predictors of grit; with other failing to find significant results from age and gender/sex. Additionally, grit does not differ by marital status in adults, and there are no significant differences observed in grit as reported by students (e.g., university, middle and high school) or according to religious background, sociocultural level, socioeconomic level, and ethnicity/race. While, some studies reveal differences between Caucasian, Hispanic, and Black students and adults.

In relation to educational variables as predictors of grit, a total of 13 studies were analyzed; 5 of these studies reveal that educational level predicts grit, while on the contrary, 4 studies do not find significant results. Furthermore, different institutional conditions and characteristics, such as location, type and climate, do not significantly impact on grit, as well as certain students’ conditions and characteristics, like, for example, enrollment status, specialties, or credit hours earned. However, negative effects on students’ grit levels are found for those who self-identified as international students, those who get a high GPA and those attending charter schools.

In relation to professional variables, 5 studies were analyzed with no statistically significant differences on grit between the different occupational types and professional performance, like, for example, teacher trust in the principal and teacher efficacy.

Finally, in relation to personal variables as predictors of grit, 28 studies were analyzed. Non-cognitive factors, like passion, dispositional resilience, sense of relatedness to various social partners (e.g., friends, teachers and parents), orientations to happiness, optimism, self-control, and positive affect have impact on grit, in the same way as cognitive factors, such as cognitive ability, general intelligence, perceived stress, hope, search for meaning, and beliefs about intelligence. Moreover, self-transcendent motives, intrinsic self-oriented motives, extrinsic self-oriented motives for going to college, self-efficacy, purpose commitment, and achievement goal orientations significantly predict grit levels. Exercisers among university students, staff, and faculty report significantly higher grit levels than non-exercisers, dedication, year and reason to running among runners predicts grit, and the fractional amplitude of low-frequency fluctuations in the right dorsomedial prefrontal cortex (fALFF) could significantly predict individual differences in grit, and partially mediates the influence of grit on academic performance. However, family support does not contribute to the overall prediction of grit among Latina/o college students, and there is no relationship between students’ placements along Perry’s Scheme of Intellectual and Ethical Development (i.e., how students understand the world and the nature of knowledge) and grit. Interventions aimed at developing solutions to achieve a goal (grit as either primary or secondary dependent variable), as a brief mindfulness and yoga intervention, a parent training and cognitive training for treating attention-deficit/hyperactivity disorder, a college and career coaching program, a youth mentoring partnership’s friend fitness program, and a grit lesson program for parent and teachers do not show impact on grit. Conversely, a grit lesson program for students reveals impact on grit.
Although there is consensus among the main initiatives (e.g., What Works Clearinghouse [WWC], Best Evidence Encyclopedia) to produce syntheses of evidences of effectiveness based on meeting rigorous standards of evaluation, and showing positive impacts, the fact is, as Slavin (2016) points out, that deciding the impact and “weight” of those methodological features (i.e., effect size, sample size, research-made vs. independent measures, publication bias, or interactions among other factors) on program effectiveness bring along important differences. Those differences could determine, in turn, whether that level of evidence is rated as “strong” (i.e., supported by at least one randomized study), “moderate” (i.e., supported by at least one quasi-experimental study), or “promising” (i.e., supported at least by one correlational study with pretests as covariates), if we follow, for example, the categories proposed by the Every Student Succeeds Act in the United States of America.

In that regard, a closer look to the data found shows that the current capacity of grit to predict educational success across domains (e.g., educational attainment, educational performance) or target populations (e.g., school and university students, general populations) rest on two sets of evidences. One set of evidences could be rated as “strong” based on the high-quality evaluation research designs supporting those evidences, and the other one could be rated as “promising”, because of the lower quality of the research design of the studies accounting for those evidences.

In the first case the strength of the evidences is limited, however, by two factors: (a) the scarce number of high-quality studies found depending on whether we follow WWC standards and categories (i.e., cognitive skills), and (b) the existence of studies reporting the same outcome measure making vs. independent measures, publication bias, or interactions among other factors) on program effectiveness bring along important differences. Those differences could determine, in turn, whether that level of evidence is rated as “strong” (i.e., supported by at least one randomized study), “moderate” (i.e., supported by at least one quasi-experimental study), or “promising” (i.e., supported at least by one correlational study with pretests as covariates), if we follow, for example, the categories proposed by the Every Student Succeeds Act in the United States of America.

In that regard, a closer look to the data found shows that the current capacity of grit to predict educational success across domains (e.g., educational attainment, educational performance) or target populations (e.g., school and university students, general populations) rest on two sets of evidences. One set of evidences could be rated as “strong” based on the high-quality evaluation research designs supporting those evidences, and the other one could be rated as “promising”, because of the lower quality of the research design of the studies accounting for those evidences.

In the first case the strength of the evidences is limited, however, by two factors: (a) the scarce number of high-quality studies found (i.e., Gerhards & Gravert, 2015; Lucas et al., 2015; Sending, 2014), which reduces the necessary statistical power of studies supporting grit as predictor of personal outcomes (i.e., cognitive skills), and (b) the existence of studies reporting the same outcome measure but failing to reach statistical significance (i.e., Gerhards & Gravert, 2015; Lucas et al., 2015; Sending, 2014). In any case, these results yield concern about the validity of the strength found and doubts on whether these evidences qualify or not as “acceptable” or “strong”, depending on whether we follow WWC standards and categories (i.e., programs can be considered effective if they do not analyze data at the proper level but produce an effect size of +0.25 or more) (WWC,

---

Table 2. Studies Using Grit as Outcome

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Statistically not significant results</th>
<th>Statistically significant results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socio-demographic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender/sex</td>
<td>Arouty, 2015; Clark, 2016; Cross, 2013; Duckworth and Quinn, 2009; Egalite et al., 2016; Eskreis-Winkler, Shulman, Beal, et al., 2014; Johnson, 2014; Sending, 2014; Stewart, 2012; Sutton et al., 2014; Verdasco, 2016; Walker et al., 2016; Waring, 2015; Weston, 2015</td>
<td>Cross, 2013; Hill et al., 2014; Von Culin et al., 2014; Yeager et al., 2014; Zimmermann and Brogan, 2015</td>
</tr>
<tr>
<td>Ethnicity/race</td>
<td>Arouty, 2015; Bates, 2011; Egalite et al., 2016; Richmond, 2015; Waring, 2015; Weston, 2015</td>
<td>Johnson, 2014; Yeager et al., 2014; Verdesco, 2016</td>
</tr>
<tr>
<td>Socioeconomic/cultural level</td>
<td>Arouty, 2015; Mills, 2015; Weston, 2015</td>
<td>Arouty, 2015; Mills, 2015; Weston, 2015</td>
</tr>
<tr>
<td>Religious background</td>
<td>Egalite et al., 2016</td>
<td></td>
</tr>
<tr>
<td>Marital/relationship status</td>
<td>Waring, 2015</td>
<td></td>
</tr>
<tr>
<td>Educational</td>
<td>Cross, 2013; Fillmore, 2015; Lerner, 2013; Salles et al., 2014</td>
<td>Egalite et al., 2016; Scholl et al., 2014; Vakil, 2014; Verdesco, 2016; Weston, 2015</td>
</tr>
<tr>
<td>GPA</td>
<td>Sutton et al., 2014</td>
<td>Fillmore, 2015</td>
</tr>
<tr>
<td>Institutional conditions and characteristics</td>
<td>Dobbie and Fryer, 2014; Incantalupo-Kuhn, 2015; Johnson, 2014</td>
<td>West et al., 2014</td>
</tr>
<tr>
<td>Students conditions and characteristics</td>
<td>Johnson, 2014; Salles et al., 2014</td>
<td>Scholl et al., 2014</td>
</tr>
<tr>
<td>Professional</td>
<td>Aycok, 2013; Kuhner, 2015; Kraft and Grace, 2015; Madden, 2015</td>
<td></td>
</tr>
<tr>
<td>Performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupational Type</td>
<td>Cross, 2013; Fillmore, 2015; Lerner, 2013; Salles et al., 2014</td>
<td>Egalite et al., 2016; Scholl et al., 2014; Vakil, 2014; Verdesco, 2016; Weston, 2015</td>
</tr>
<tr>
<td>Personality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-cognitive factors</td>
<td>Goodman et al., 2014; Karlson, 2014; Moore, 2015</td>
<td>Chambers et al., 2012; Datu, 2017; Hill et al., 2014; Incantalupo-Kuhn, 2015; Kench et al., 2016; Lyon, 2014; Moshier et al., 2016; Slack, 2014; Von Culin et al., 2014</td>
</tr>
<tr>
<td>Cognitive factors</td>
<td>Rutledge, 2012;</td>
<td>Fillmore, 2015; Lyon, 2014; Richmond, 2015; Saunders-Scott et al., 2014; Vela et al., 2015; Yeager et al., 2014</td>
</tr>
<tr>
<td>Motivation</td>
<td>Anderson, 2012</td>
<td>Akin and Asljan, 2014; Hill et al., 2014</td>
</tr>
<tr>
<td>Physical activity</td>
<td>Major, 2013;</td>
<td>Madrigal, 2014; Reed et al., 2014; Vakil, 2014</td>
</tr>
<tr>
<td>fALFF</td>
<td>Vela et al., 2015</td>
<td>Wang et al., 2017</td>
</tr>
</tbody>
</table>

Note. Numbers represent studies references.
1Study 1; 2Study 2; 3Study 3; 4Study 4; 5Study 5; 6Study 6. Interventions aimed at developing solutions to achieve a goal (grit as either primary or secondary dependent variable). 7Grit as a part of model.

Discussion

According to our initial research objective and questions, our results suggest, firstly, that current scientific investigation focused on this construct is not as systematic and as rigorous, as we would like to or, in other words, that there is a need to strengthen the quality of the evaluation research designs utilized, as well as to carry out more impact evaluation studies on grit as predictor and as outcome. Secondly, our results reveal that there are evidences of effectiveness of grit as predictor of educational, professional, and personal success, and also that grit can be influenced by an array of socio-demographic, educational, professional, and personal factors. However, the limited number of studies found bearing empirical evidences based on credible research designs, as well as the relative strength of those evidences of effectiveness, prevent us from drawing causal links on grit as predictor of the different types of success analyzed, as well as grit as outcome.

Although there is consensus among the main initiatives (e.g., What Works Clearinghouse [WWC], Best Evidence Encyclopedia) to produce syntheses of evidences of effectiveness based on meeting rigorous standards of evaluation, and showing positive impacts, the fact is, as Slavin (2016) points out, that deciding the impact and “weight” of those methodological features (i.e., effect size, sample size, research-made vs. independent measures, publication bias, or interactions among other factors) on program effectiveness bring along important differences. Those differences could determine, in turn, whether that level of evidence is rated as “strong” (i.e., supported by at least one randomized study), “moderate” (i.e., supported by at least one quasi-experimental study), or “promising” (i.e., supported at least by one correlational study with pretests as covariates), if we follow, for example, the categories proposed by the Every Student Succeeds Act in the United States of America.

In that regard, a closer look to the data found shows that the current capacity of grit to predict educational success across domains (e.g., educational attainment, educational performance) or target populations (e.g., school and university students, general populations) rest on two sets of evidences. One set of evidences could be rated as “strong” based on the high-quality evaluation research designs supporting those evidences, and the other one could be rated as “promising”, because of the lower quality of the research design of the studies accounting for those evidences.

In the first case the strength of the evidences is limited, however, by two factors: (a) the scarce number of high-quality studies found (i.e., Gerhards & Gravert, 2015; Lucas et al., 2015; Sending, 2014), which reduces the necessary statistical power of studies supporting grit as predictor of personal outcomes (i.e., cognitive skills), and (b) the existence of studies reporting the same outcome measure but failing to reach statistical significance (i.e., Gerhards & Gravert, 2015; Lucas et al., 2015; Sending, 2014). In any case, these results yield concern about the validity of the strength found and doubts on whether these evidences qualify or not as “acceptable” or “strong”, depending on whether we follow WWC standards and categories (i.e., programs can be considered effective if they do not analyze data at the proper level but produce an effect size of +0.25 or more) (WWC,
et al. (2007), Eskreis-Winkler, Shulman, Beal et al. (2014), and Von Sending, 20143) bearing significant statistical results, although for (i.e., performance) variables, and another high-quality study (i.e., reporting no statistically significant impact on grit from professional university records and databases (e.g., Richmond, 2015) or self-report within different subgroups of the sample (e.g., Weston, 2015), (c) the use of researcher-made vs. independent measures (e.g., Van Doorn, 2014), (d) the lack of experimental-control comparison with evidence of no pre-test differences among the two groups (e.g., Goodman et al., 2014), (e) how students records were collected via school and university records and databases (e.g., Richmond, 2015) or self-report (e.g., Cross, 2013), (f) which subjects are analyzed, since the hypothesis that grit would interact with previous grades to predict current grades was not supported (e.g., Hatchmonji, 2016), (g) whether those records correspond to specific subjects (e.g., Fillmore, 2015) or the total or average of those subjects (e.g., Zimmerman & Brogan, 2015), (h) whether they refer to semesters (e.g., Engel, 2013), academic or school year (e.g., West et al., 2014), or accumulated (cumulative college GPA; final law school GPA) (e.g., Wang et al., 2017), and (i) what interactions are analyzed (e.g., grit data against grades from the earliest years of college could not indicate actual use of grit behaviors, as students are only beginning their long-term goal of obtaining a bachelor's degree is reported by Black, 2014). In this context, the identification of studies showing no significant results with small sample size (e.g., Salles et al., 2014), or the way “performance” is measured (e.g., study 39 measured “task performance” throughout a questionnaire including a selection of items adapted from other scales) represent examples, as Slavin (2016) points, of how studies methodological factors can vary and their potential effects on the outcomes measured as effectiveness (i.e., effect size).

Additionally, some of the statistically significant results identified are modest in terms of variance explained (e.g., Muenks et al., 2016), which aligns with that of Rimfeld et al. (2016), in which grit added only 0.5% to the prediction of another educational variable like the General Certificate of Secondary Education exams. Furthermore, according to these authors, the small effect coming from that high-quality research study (using a large representative sample of the United Kingdom population at the age 16) warrant concern against the decision of educational policies directives in the United States and the United Kingdom. On the contrary, our results partially endorse the decision of promoting grit as a critical factor for success in the 21st century, as reported by Shechtman et al. (2013). Moreover, our results on grit studies as significant predictor of life outcomes, when controlling for Big Five personality factors, although explaining only minor incremental variance (e.g., Duckworth & Quinn, 2009), align with those of Duckworth and Eskreis-Winkler (2013), Duckworth et al. (2007), Eskreis-Winkler, Shulman, Beal et al. (2014), and Von Culin et al. (2014) and endorse the salient capacity of grit as unique psychological factor determining success.

The situation underlying grit research as outcome is more complex, with just one high-quality study (i.e., Kraft & Grace, 2015) reporting no statistically significant impact on grit from professional (i.e., performance) variables, and another high-quality study (i.e., Sending, 2014) bearing significant statistical results, although for two key socio-demographic variables (i.e., age and gender/sex as predictor). In the literature, grit was found to increase over the lifespan (Duckworth et al., 2007), so the apparently contradictory results on this variable could be attributed to particular methodological bias or a combination of those like the presence of a sub sample within the sample (i.e., Cross, 2013). Similar controversial results were found for other predictors like “gender/sex” and “ethnicity/race”, although we have failed to find more plausible answers than the dominant presence of white female samples. The rest of the studies adopting pre-experimental, ex-post facto, or correlational designs on socio-demographic factors (i.e., gender/sex), educational factors (i.e., GPA), and personal factors (i.e., non-cognitive factors, cognitive factors, and physical activity) report no significant differences (i.e., Sutton et al., 2014) and significant differences (i.e., Lyon, 2014) respectively. Additionally, the variability of variables measured under the category named “non-cognitive factors” (e.g., “resilience”, “perceived stress”, “hope”, “higher levels of theoretically relevant masculine norms”, or “openness to experience”), and the identification of studies with apparently contradictory results (e.g., Incantalupo-Kuhner, 2015) recommends to be cautious when interpreting these results, because of the limitations of current available measures accompanying some of those variables (Duckworth & Yeager, 2015), as well as the potential methodological bias underlying such results due to the adoption of less reliable research designs (i.e., correlational). Moreover, the category named “cognitive factors” representing variables like “cognitive ability” (i.e., Yeager et al., 2014), “implicit theories of intelligence” (i.e., Fillmore, 2015), “perceived stress” (i.e., Saunders-Scott et al., 2014), or “subjective happiness” (i.e., Vela et al., 2015) needs more research to improve the predictive usefulness of some of those concepts like “conscientiousness”, as some studies suggest (Credé et al., 2016; Duckworth et al., 2007; Eisenberg et al., 2014; MacCann et al., 2009).

In sum, from the evidence-based movement perspective, our synthesis reveals that, on one hand, the research field on grit as predictor of educational, professional, and personal success and grit as outcome needs to increase the robustness of the evidences found as well as their comparability. On the other hand, the field is moving fast in the “pipeline” classification suggested by the Institute of Education Sciences and National Science Foundation (2013). Indeed, our results on grit as predictor and particularly as outcome conform to that classification of research development with abundance of studies corresponding to the first two types of studies (i.e., “foundational” and “early-stage or exploratory research”), which examines relationships (usually correlational rather than causal) among relevant constructs in learning and education to provide the fundamental knowledge contributing to set the theoretical (i.e., Duckworth, 2006; Duckworth et al., 2007) and methodological bases (i.e., Duckworth & Quinn, 2009) even internationally (i.e., Akin et al., 2011; Nishikawa et al., 2015) to inform, guide, support, and conduct more causal research studies in the following years. Furthermore, our results also reveal that there is a growing body of more evolved research studies corresponding to the next level category named “design and development” (type #3), focused on grit as predictor or grit as outcome (i.e., Lyon, 2014) successfully contributing to achieve the goal of improving grit levels across populations intentionally. Although, as Duckworth and Yeager (2015) recommend, it is more important to start working on reaching consensus about the specific attributes to focus on and, in particular, their definition and measurement.

Hopefully, in the near future, the progressive implementation of more advanced types of research like “efficacy research” (type #4), “effectiveness research” (type #5) and, eventually, “scale-up research” (type #6) will allow researchers, educators, and practitioners to distill more effective strategies and interventions on grit in a wider range of populations and settings. Concurrently, complementary research strategies aimed at expanding large-scale longitudinal data on different subgroups of populations appear as unavoidable, if we want to accelerate the pace of the transference of the evidence-based movement and their benefits into the field of education in general and of grit in particular. In such scenario, we will be able to translate more effectively our educational and training efforts on grit into better supply of skilled, employable youth as Hossain and Bloom (2015), and Schleicher (2018) recommend.
Limitations

The protocol we developed following the guidelines provided by The Campbell Collaboration (2015) was compatible with the realization of a systematic review and a meta-analysis. However, as we completed the process of extracting the prescribed data from the studies sampled, we realized that a meta-analysis would not be feasible. The main reasons are related to underreporting problems affecting some of the quantities (i.e., alpha, power, effect size, N) and qualities (i.e., the hypothesis/aim/objective of the study is clearly described, measures are accurate – valid and reliable –, randomization is clearly stated) of most of the studies sampled.

Our final review intends to be as valid, unbiased, and meaningful as possible to respond to our broad research questions. However, the process of producing high-quality syntheses of program evaluation research can be compromised by subjectivity permeating supposed objective decisions and method (e.g., the framing of the question, the choice of eligible studies, the selection of comparisons, populations, and outcomes of interest, the types of data extracted), which allows for substantial diversity in the final results (Ioannidis, 2016). Thus, the design and implementation of the research protocol has unveiled several potential sources of bias. For example, the fact that grit is a research topic “under construction” has conditioned the ‘objective’ inclusion criteria for systematic reviews recommended by Slavin (2008), Van Klaveren and De Wolf (2013), or WWC (2014). In this vein, considering that we pursued to portraying the state of the art concerning the quality of grit research, we have faced the typical reviewers’ dilemma about how much stringent the inclusion criteria should be, because of the risk of finding a few studies (or even a single study) qualifying as examples of those (high or strong) quality level of evidences. The broader the inclusion criteria, the larger the number of studies sampled, although at the expense of having a very wide range in study quality (i.e., research design), which makes more difficult to determine unique effects of factors that affect outcomes even in well-controlled studies (Cheung & Slavin, 2016). That is the case of this review, in which the research questions posed (i.e., state of the art on grit research) has lead, in spite of using only Duckworth’s (2006) definition, on one hand, to a pull of studies with an enormous variability of methodological features (e.g., researcher-made vs. independent measures, sample size, research design, lack of information on correlations among factors) and, on the other hand, to a too small number of high-quality studies for adequate statistical power. Consequently, both features have prevented this review, in turn, from applying an evaluation synthesis method to rate the strength of the evidence base and the size of the anticipated effects for each program in an easy and well-justified manner, as Slavin (2008) recommends.

Additionally, the prevalence of reporting bias affecting social sciences literature in general identified by several publications (American Psychological Association, 2008; Higgins et al., 2011; The Campbell Collaboration, 2015) applies also to this body of research. In fact, the reviewer’s inferences about the studies methodological features because of the underreporting problems affecting the majority of the studies reviewed, particularly those with correlational or cross-sectional design, recommends to take these results with caution. Additionally, the lack of previous syntheses research on this particular topic or subject, including the relationship between methodology and effect size, makes it impossible to compare our results with others as recommended (Slavin, 2016).

Conclusions

Embracing evidences as a basis for practice requires the existence of scientifically valid and readily interpretable syntheses of research on a particular practice or program and our findings suggest that the implementation of full-scale programs to enhance grit might be premature. However, using the Education Sciences and National Science Foundation framework as a reference, we found evidence of types one through three studies regarding the grit construct. Based on these findings, research on the construct is off to a good start, even though more research is needed. For example, the results on educational variables as predictors of grit, apart from being apparently contradictory, do not yield much munition to guide educational interventions or programs. Alternatively, our data on interventions aiming to modify personal factors do provide an array of promising interventions to teach and foster qualities related to grit (e.g., optimism, self-control, perseverance) via educational interventions targeting not only youth but also parents and teachers, as Lyon (2014) demonstrates.

In this line, more high-quality evaluation and research designs are necessary to improve the primary research evaluation and reporting standards in both scenarios, which will lead towards more proven practices and programs on grit development and to learn which specific activities are most effective in producing better outcomes. As Ioannidis (2016) points, “more should be done to improve the quality of the primary evidences rather than expect systematic reviews to correct deficiencies after the fact” (p. 504).

Finally, aligned with Glennerster (2012) arguments on the power of evidence, school-to-work legislation has to be developed to ensure that intervention programs and training activities on grit in-and-out-of-school time could contribute effectively to improve the educational and personal outcomes of children and young adults.

Conflict of Interest

The authors of this article declare no conflict of interest.

References

References with an asterisk refer to articles included in this SR.


