Motivation Interventions in Education: 
A Meta-Analytic Review

Rory A. Lazowski
James Madison University

Chris S. Hulleman
University of Virginia

This meta-analysis provides an extensive and organized summary of intervention studies in education that are grounded in motivation theory. We identified 74 published and unpublished papers that experimentally manipulated an independent variable and measured an authentic educational outcome within an ecologically valid educational context. Our analyses included 92 independent effect sizes with 38,377 participants. Our results indicated that interventions were generally effective, with an average mean effect size of $d = 0.49$ (95% confidence interval = [0.43, 0.56]). Although there were descriptive differences in the effect sizes across several moderator variables considered in our analyses, the only significant difference found was for the type of experimental design, with randomized designs having smaller effect sizes than quasi-experimental designs. This work illustrates the extent to which interventions and accompanying theories have been tested via experimental methods and provides information about appropriate next steps in developing and testing effective motivation interventions in education.

Keywords: meta-analysis, motivation, interventions, education, research synthesis

Over the past two decades, intervention research in the field of education has been on the decline (Hsieh et al., 2005; Robinson, Levin, Thomas, Pituch, & Vaughn, 2007). Motivation research is no exception. Despite the considerable volume of theoretical, qualitative, observational, and correlational studies, there have been few experimental tests of motivation theory in the field of education (Wentzel & Wigfield, 2007). This trend has persisted despite calls for increasing intervention and use-inspired research (e.g., Blackwell, Trzesniewski, & Dweck, 2007; Hidi & Harackiewicz, 2000; Maehr & Meyer, 1997; Midgley & Edelin, 1998; Pintrich, 2003; Wentzel & Wigfield, 2007). Although observational and correlational research can generate and test hypotheses, intervention research...
(i.e., empirical investigations that manipulate an independent variable) provides valuable information about what happens when we attempt to enhance educational outcomes through intentional manipulation. From a theoretical perspective, intervention studies help move the field forward by providing insight about the causal relationships between motivation constructs and educational outcomes, or between educational settings and motivation outcomes (Shadish, Cook, & Campbell, 2002; Tunnell, 1977). Because interventions represent the operationalized theory in action, they provide a strong test of the theory as applied in an educational context. From a practical perspective, intervention studies facilitate our understanding about which interventions are most effective in improving educational outcomes in a way that observational research cannot. This understanding can guide recommendations for educational practice based on appropriate scientific evidence.1

The purpose of the present investigation was to conduct a meta-analytic review of educational intervention studies grounded in motivation theory. Our goal is to showcase the work on social psychological interventions in education that leverage motivational processes to enhance student learning outcomes. Meta-analysis is a tool that enables researchers to systematically quantify research in a field. Although narrative reviews exist (e.g., Martin & Dowson, 2009; Yeager & Walton, 2011), to our knowledge a meta-analysis examining motivation interventions conducted in field settings has not yet been conducted. Our goal was to identify theoretically grounded motivation interventions that had been experimentally tested in educational contexts, and examine the extent to which these interventions affected student outcomes. We also hoped to explain variability in intervention effectiveness across study characteristics, such as theoretical framework, experimental design, and grade-level of participants. As a result, this study synthesizes the available empirical evidence regarding motivation interventions in education, and highlights the appropriate next steps in developing, testing, and implementing effective motivation interventions in educational contexts.

What Is Motivation?

The Latin derivative of motivation means “to move.” In lay terms, motivation is often considered to be either intrinsic (from internal sources, such as the pure enjoyment of task engagement) or extrinsic (from external sources, such as receiving financial compensation). However, for those studying motivation in psychology and education, there are a multitude of types and qualities of motivation, such as needs, drives, goals, aspirations, interests, and affects, to name a few. The diverse nature of motivation and associated processes has been studied from multiple perspectives in psychology—cognitive, developmental, educational, social—and generated an extensive list of constructs and theoretical frameworks, resulting in an extensive research literature (for reviews, see Brophy, 2004; Perry, Turner, & Meyer, 2006; Pintrich, 2003; Schunk, Pintrich, & Meece, 2008; Stipek, 2002; Wigfield & Eccles, 2002). In general, theories of motivation are concerned with the energization and direction of behavior (Pintrich, 2003). In education, these theories tend to use a social–cognitive foundation, which emphasizes an individual’s perception of her/himself within the social context (Perry et al., 2006), instead of relying on more biologically based drives and tension systems.
Motivation Interventions

(Weiner, 1980). Social—cognitive theories of motivation in social and educational psychology include those focusing on achievement motives and needs (e.g., Deci & Ryan, 1985), perceived ability and self-concept (e.g., Marsh & Shavelson, 1985), perceived value for and interest in an activity (e.g., Eccles et al., 1983; Hidi & Renninger, 2006), goals (e.g., Gollwitzer, 1999; Nicholls, 1984), attributions about success and failure (e.g., Weiner, 1980), emotions (e.g., Pekrun, 2006), and potential future identities (e.g., Markus & Nurius, 1986).

Navigating this dizzying array of constructs can be a challenge even for the seasoned researcher. In education, Schunk, Pintrich, and Meece (2013) are on their fourth edition of a textbook that summarizes, but does not synthesize, motivation constructs and theories relevant to education. Although there may be a need for a synthetic theory of motivation in education, this task was too big even for Schunk et al.’s book, so we will not attempt it here. Instead, our focus is to synthesize findings relevant to interventions that target motivational processes. Although our definition is broad—it includes processes that are affective, behavioral, and/or cognitive—the inclusion criteria for entry into our analyses is precise. As explained in detail later, the interventions in our review must have been grounded in motivation theory and designed to leverage motivational processes as their primary aim.

Why a Meta-Analysis on Motivation Interventions?

Given our previous argument, it is pertinent to wonder why we chose motivation interventions for this meta-analytic review. After all, if the topic of motivation is too big even for a book, then what could we hope to accomplish in a journal article? Our decision was driven by two important factors. First, a rich body of research in educational and social psychology has clearly demonstrated that student motivation is essential for learning, and if left unguarded declines in motivation will undermine system effectiveness. Not only do motivated students learn more, they persist longer, produce higher quality work, and score higher on standardized achievement tests, particularly if they are motivated by relatively intrinsic, as compared with extrinsic, reasons (e.g., Fredericks et al., 2011; Maehr & Midgley, 1999; National Research Council & Institute of Medicine, 2004). For example, in their meta-analysis of 109 studies, Robbins et al. (2004) found that measures of student motivation were predictive of academic performance and persistence in college, after controlling for prior measures of performance and academic ability. Casillas et al. (2012) found that motivation in middle school accounted for as much variance in high school GPA as middle school grades did for the nearly 4,700 students in their study.

Unfortunately, declines in student motivation are a systemic problem in schools and threaten educational equity (Nicholls, 1979). Decreases in student motivation span grade-levels (from elementary to high school) and types of motivation. For example, Lepper, Corpus, and Iyengar (2005) demonstrated that both intrinsic and extrinsic motivation declined from 3rd to 8th grade, and Jacobs, Lanza, Osgood, Eccles, and Wigfield (2002) revealed that student’s perceived competence in, and value for, math and English declined from 1st to 12th grade. In a recent poll of elementary through high school students (Gallup, Inc., 2014), 8 of 10 reported to be “learning with a positive emotional tone and persevering in the face of
challenges.” This number drops to 6 of 10 in middle school and 4 of 10 in high school. In a national survey of high school dropouts, 69% reported that their schools failed to motivate them (Bridgeland, DiLulio, & Morison, 2006).

The conclusion is that although motivation is critical for learning, we do not seem to be enhancing, or even maintaining, motivation to learn in school at a systematic level. How can we address this critical, fundamental issue in education? This leads us to our second reason for doing this meta-analysis: What do we know about how to enhance motivation and learning in school? Fortunately, recent narrative reviews reveal numerous brief, social—psychological interventions, designed to leverage motivational processes that increase student learning outcomes. For example, in their narrative review of social—psychological interventions in education, Yeager and Walton (2011) provide a conceptual frame for understanding how and why these seemingly small interventions (in comparison with comprehensive school reform) could have such powerful impacts on student learning outcomes.

The purpose of our review is complementary, yet distinct, from Yeager and Walton’s (2011) paper. Rather than using interventions as examples of how psychological processes can instigate change within a system, as Yeager and Walton did, we systematically examined the landscape of motivation intervention research to estimate the average effect size of such interventions on student learning outcomes as well as some potential moderating variables. Our goal is to highlight the power of social psychological, motivational interventions in education by quantifying their effect size and robustness across outcomes and grade levels. When combined with the theoretical framework for understanding these effects provided by Yeager and Walton, we hope to encourage researchers, program developers, and practitioners to continue developing and testing educational interventions that can make an impact on educational practice.

**Interventions and Motivation Theory**

Education researchers are ultimately interested in how to structure the educational context to maximize student learning outcomes. In other words, we aim to develop an intervention, or interventions, that facilitate student learning and performance. This requires testing the extent to which the interventions, based on our ideas, created the kind of change in students and teachers that we had envisioned. If not, we can go back to the drawing board to revise the intervention, our theories, or both. Without this kind of idea testing, our theories will not be pushed to grow, and knowledge about how to best structure educational environments will be limited. Incorporating intervention studies into how we think about our theories is aligned with movements in other fields to more quickly translate research findings into practice (Daniel, 2012), such as improvement science efforts in health care (e.g., Berwick, 2008) and education (e.g., Bryk, Gomez, Grunow, & LeMahieu, 2015). When conducted as a part of a complete methodological approach that includes observations and interviews (see Brown, 1992; Design-based Research Collective, 2003; Harackiewicz & Barron, 2004; Harackiewicz & Hulleman, 2010; Hulleman & Barron, 2016), intervention studies offer the opportunity to make great advances in our theoretical and practical knowledge about education.
Our review of the research literature focused on interventions designed to enhance student motivation and achievement that were implemented within an ecologically valid educational context. We defined an intervention as a manipulation implemented by an external agent (i.e., teacher, researcher) that was intended to change students’ cognitions, emotions, and/or behaviors. We were interested in intervention studies that used some type of control group (i.e., matched control group, randomization to conditions) that bolstered internal validity (i.e., claims for inferring cause and effect; Shadish et al., 2002). To operationalize an ecologically valid educational context, we used Tunnell’s (1977) three dimensions of naturalness: natural treatments are naturally occurring events to which the participant is exposed (e.g., pedagogical practices, curriculum); natural settings are those that are not perceived to be established for the purposes of research (e.g., almost any setting outside the laboratory; see Shadish et al., 2002); and, natural behavior occurs on its own without experimental intervention (e.g., statewide mandated standardized tests). Because we are interested in how motivation theory has been tested through interventions that have direct applications to educational practice, our review included studies that contained one or more dimensions of naturalness. Intervention studies that contain these dimensions of naturalness are more likely, although not guaranteed, to have results that will generalize to other settings (i.e., external validity).

Although not representative of all motivation theories, the studies comprising our review come from some of the most dominant, contemporary theoretical perspectives in student motivation. These theories include achievement emotions, achievement goals, attribution, expectancy-value, goal setting, implicit theories of intelligence (mindsets), interest, need for achievement, possible selves, self-affirmation, self-confrontation, self-determination, self-efficacy, social belongingness, and transformative experience. These theoretical frameworks have produced a variety of different interventions that target student motivation as an instigator of enhanced learning outcomes. Below, we offer a brief overview of the 15 theoretical frameworks and select examples of interventions. A more detailed description of each theoretical framework and associated interventions can be found in Supplementary Tables S1 and S2 in the online version of the journal.

Theories of achievement emotions have produced interventions that reduce worries about test taking through expressive writing (e.g., Ramirez & Beilock, 2011), or that enhance happiness and well-being through personal improvement exercises (e.g., Fordyce, 1977, 1983). For example, Jamieson, Mendes, Blackstock, and Schmader (2010) randomly assigned undergraduates who were preparing to take the Graduate Record Examination (GRE) to either an anxiety reappraisal condition or a control group. Prior to taking a practice GRE, students in the reappraisal condition read information that explained that feeling anxious while taking standardized tests was not only normal, but that research showed that this arousal was not detrimental and could in fact improve performance on the tests. They were also instructed that if they did feel anxious, they should remind themselves that this anxiety could actually be helping them do well on the test. Students in the control condition read instructions only that feeling anxious during standardized tests was a normal response. When taking the actual GRE 1 to 3 months later, students in the reappraisal condition performed better than students in the control condition.

Motivation Interventions
Achievement goal theory has produced interventions that encourage students to adopt mastery goals for learning and to value the process of learning as much as the product (e.g., Muis, Ranellucci, Franco, & Crippen, 2013). For example, Hoyert and O’Dell (2006) conducted two intervention studies aimed at altering achievement goal orientations among struggling college students. In the first study, a guest speaker came to the class and provided a lecture and discussion about how to set mastery goals and to examine various meanings behind failure. Students then completed exercises designed to influence their adoption of mastery goals covering topics such as goal setting and study strategies. The writing assignment required students to write about defining goals, describe individuals who illustrate the traits associated with the goal type, and consider personal experiences related to the goal orientation. In the second study, the intervention was delivered via a computerized program. The results of both studies showed that students in the experimental condition showed increases in grades and mastery goal orientation compared with students in the control condition.

Attribution theory has produced interventions that help students attribute academic struggles to lack of effort through individual and group activities (e.g., Hall et al., 2007; Hall, Hladkyj, Perry, & Ruthig, 2004). In a series of studies, Wilson and Linville (1982, 1985) tested the effects of an attribution intervention on academic performance. In the original study (1982), students were randomly assigned to either an experimental or control condition. Students in the experimental condition watched videotapes and were shown statistics indicating how students typically struggled academically during their freshman year but improved afterwards. Students in the control condition did not receive this information. The results indicated that students in the attribution retraining group, compared with those in the control group, increased their GPA from before to after the study and were less likely to drop out from school the following year.

The expectancy-value framework has inspired interventions that help students discover the relevance of academic material to their lives through writing exercises (e.g., Hulleman & Harackiewicz, 2009). In a series of studies, Hulleman and colleagues have tested a writing intervention that asks students to write about how some aspect of the course material relates to their lives in some way. In one study, Hulleman and Harackiewicz randomly assigned high school science students to write about how what they were studying related to their lives, or to simply write a summary of the material they were learning. For students with low expectancies of success at the beginning of the course, students in the intervention group earned significantly higher GPAs (0.8 GPA points) than those in the control group. There were no differences for confident students.

Goal setting theory has produced interventions that help students set specific, realistic academic goals through direct instruction (Morisano, Hirsh, Peterson, Pihl, & Shore, 2010). For example, Morisano tested the effects of an online goal-setting intervention on academic achievement for struggling college students. Students randomly assigned to the experimental condition completed a comprehensive online program that asked them to identify important future goals, elaborate in detail their goals and implementation plans, and indicate how committed they were to achieve the goal(s). Students in the control group completed online tasks and wrote about topics such as positive past experiences. The results
Motivation Interventions

indicated that students in the experimental group showed significant increases in GPA as a result of the program, whereas there were no changes in GPA for students in the control group.

Dweck’s (1986) implicit theories of intelligence has inspired a series of interventions that help students realize that we can become smarter by tackling challenges (e.g., Aronson, Fried, & Good, 2002). For example, Blackwell et al. (2007) tested the effects of an in-depth intervention designed to teach seventh grade students about how the brain can become stronger through effort and challenge. Students randomly assigned to the control condition received 6 weekly, 25-minute lessons that covered the structure and function of the brain, and how the brain changes over time. Students randomly assigned to the growth mindset condition participated in two additional lessons (eight total) that focused on the incremental theory of intelligence and discussed the malleability of the brain and how learning makes students smarter. Not only were students in the growth mindset conditions rated as having increased their motivation more over the year by their teachers than those in the control group, the mindset students also increased their grades over the rest of the semester, whereas students in the control group decreased their grades.

Interest theory has produced interventions that excite and engage students in a topic through expressive writing and providing stimulating learning activities (e.g., Guthrie et al., 2006). For example, Guthrie et al. (2006) tested an intervention designed to stimulate situational interest and promote long-term interest and intrinsic motivation in the area of reading (Concept-Oriented Reading Instruction, or CORI). Third-grade students from four separate classes received science instruction that integrated fiction and nonfiction reading. The intervention involved two of the teachers delivering a higher amount of stimulating tasks to students than the other two classrooms (e.g., observations, drawings, experiments). Students in classrooms that had more stimulating tasks scored higher on reading comprehension and intrinsic motivation to read than students in the classrooms with fewer stimulating tasks.

Need for achievement theory has produced interventions that encourage students to strive for high achievement through direct instruction in understanding achievement-related emotions and cognitions, and adopting higher standards and action strategies (e.g., Cueva, 2006). In one randomized field experiment (Cueva, 2006), students participated in a series of 1-hour sessions in small groups as part of the counseling and guidance program. The 12 sessions were based on McClelland’s need for achievement (see Alschuler, Tabor, & McIntyre, 1971), and involved introducing children to achievement thoughts and emotions, and action strategies that can be used to accomplish personal goals (i.e., goal setting and monitoring of goal progress). Control participants did not receive any additional programming. The results did not reveal significant differences on either the self-reports of motivation or academic achievement.

Possible selves theory has inspired interventions that help students draw connections between successful future selves and current school involvement through interactive activities and written reflections (e.g., Oyserman, Bybee, & Terry, 2006; Oyserman, Terry, & Bybee, 2002). In a study by Oyserman et al. (2002), students randomly assigned to the experimental condition received an intervention
consisting of a 9-week after school program designed to enhance students’ abilities to see themselves as successful adults. Students completed activities that encouraged them to draw connections between these selves and their current school involvement. Students in the control condition did not receive the intervention. Results indicated that students in the experimental condition demonstrated a greater sense of bonding to school, fewer discipline referrals, and better attendance patterns compared with students in the control condition.

Self-affirmation theory has produced interventions that help students maintain self-integrity by affirming important values in writing exercises (e.g., Cohen, Garcia, Apfel, & Master, 2006). Largely, these interventions have targeted students who are most at-risk for feeling threatened in specific academic situations, such as students from groups stereotyped to underperform (e.g., women in math or science). For example, Cohen and colleagues tested a values affirmation intervention with African American and European American 7th graders from middle to lower middle class families. The experimental activity, which took about 15 minutes to complete, began by presenting all students with a list of values. Students randomly assigned to the experimental condition were instructed to choose their most important value (Study 1) or to choose two or three of their most important values (replication study). In contrast, students in the control condition were instructed to choose their least important value (Study 1) or to choose two or three of their least important values (replication study). Students in the experimental conditions wrote a passage about why their value(s) were personally important, and students in the control condition wrote about why their least important value might be important to another person. Results indicated that African American students in the experimental condition earned significantly higher grades at the end of the term compared with African American students in the control condition. No significant differences emerged in either study for European Americans.

Self-confrontation theory has produced interventions that encourage students to adopt behaviors that enable them to achieve outcomes consistent with their core values through confronting students with value discrepancies between themselves and other successful students (Greenstein, 1976). This theory suggests that motivation to change is elicited when students perceive that their behaviors and values differ from their self-conception (Rokeach, 1973). In a study by Greenstein, student teachers first completed a questionnaire where they ranked their values for teaching, and then were randomly assigned to receive feedback showing their own values, and the values of good and mediocre teachers, or not to receive feedback. The feedback showed that highly rated teachers valued mature love over a sense of accomplishment, whereas the opposite was true for mediocre teachers. Teachers who received feedback earned significantly higher scores on a behavioral measure of teaching ability than did student teachers not receiving such feedback.

Self-determination theory has inspired interventions that help students perceive more choice and control in the learning setting by helping teachers use instructional practices that support student autonomy (Reeve, Jang, Carrell, Jeon, & Barch, 2004). Many of these studies have focused on manipulating autonomy (choice and control) and intrinsic benefits as the primary drivers of the interventions. For example, Vansteenkiste, Simons, Lens, Soenens, and Matos (2005)
manipulated information that early adolescent students received when reading a text about nutrition. In one study, students were randomly assigned to receive information in a manner that primed either intrinsic (e.g., physical wellness) or extrinsic goals (e.g., physical attraction) for the task. Students were also randomly assigned to receive the information in an autonomy-supportive style (e.g., choice to follow the nutritional guidelines) or a controlling style (e.g., explicit expectations and/or pressure to follow the guidelines). The results showed that students in the intrinsic goal framing conditions scored higher in short-term and long-term conceptual learning compared with the extrinsic goal framing conditions, and students in the autonomy-supportive conditions scored higher than students in the externally controlling conditions.

*Self-efficacy* has inspired interventions that help students perceive that they can successfully complete the specific tasks and activities required for learning, such as setting and achieving realistic study goals (Bandura & Schunk, 1981). In one study (Schunk & Cox, 1986), middle school students going through six, 45-minute subtraction skill training sessions over 6 consecutive days were randomly assigned to one of three conditions. Students either received effort feedback (“You’ve been working hard”) after the third session, after the final session, or not at all. All students received performance feedback (“You’re doing OK”) throughout the sessions. The effort feedback was designed to highlight a controllable source of efficacy for students to build off of as they developed their subtraction skill. Receiving effort feedback, regardless of timing, led to increased self-efficacy, number of correctly solved problems during training, and subtraction performance at the end of the six sessions.

*Social belongingness* theory has inspired interventions that help students perceive stronger connections between themselves and important others in the learning context (Hausmann, Ye, Schofield, & Woods, 2009) and to reduce feelings of uncertainty about belonging (Walton & Cohen, 2011). In the Walton study, students randomly assigned to the experimental condition were first provided information suggesting that most college students experience some sense of worry or doubt about belonging on campus but that these feelings diminish over time, and then developed and delivered a presentation to future students with a hopeful message about belonging. Students in the control condition were provided information suggesting college students’ social–political beliefs become more developed over time. For students most likely to feel uncertain about belonging, in this case African American students, those in the experimental condition reported higher levels of academic fit immediately after the intervention and larger increases in GPA from freshmen to senior year, compared with students in the control condition.

The model of *transformative experience* has inspired interventions that encourage teachers to reframe learning in a way that enhances value of the content to students’ everyday experience through various instructional strategies (Pugh, 2014). In one study (Pugh, 2011), one teacher was instructed in the Teaching for Transformative Experience in Science model, which focused on three principles to scaffold learning: frame the content as ideas to be imagined about rather than as concepts to be learned, re-seeing objects as new ideas, and modeling transformative experience. A control teacher in the same school was not exposed to these.
principles. Students in the experimental class scored higher on self-reports of transformative experiences and on an assessment of science knowledge than those in the control class.

**Current Research**

This meta-analytic review is focused on published and unpublished studies of field interventions in the area of motivation in educational settings. For the purposes of this review, we define an intervention study as an empirical investigation that manipulated an independent variable and compared the experimental group with some type of control group. Motivation interventions were those that directly targeted the energization and direction of behavior. The studies comprising the review come from many of the most dominant, contemporary theoretical perspectives in educational, social, and developmental psychology.

**Method**

**Search Criteria**

The literature review for published studies was conducted using a variety of internet search engines. One of the primary search tools was the APA PsycNET Platform. This platform included PsycINFO, PsycARTICLES, PsycBOOKS, PsycCRITIQUES, and PsycEXTRA. Other search engines included ERIC (EBSCO), Education Search Complete (EBSCO), and Google Scholar. We also cross-referenced from articles found in the search. We did not limit our search to any specified span of time or range of dates; therefore, we included all years in the search. The final search for studies was conducted in May 2015.

We also searched for “grey” literature in an effort to limit selection and publication bias. Grey literature represents studies that have not been published in sources such as books or journals. According to recent reviews, published studies tend to show greater treatment effects than those found in the grey literature (Hopewell, McDonald, Clarke, & Egger, 2007). We attempted several search strategies to uncover grey literature. In addition to search engines that include grey literature (e.g., ERIC), we also searched ProQuest Dissertations and Theses, contacted prominent researchers in the field who conduct motivation intervention research (including a mass email to the AERA Motivation in Education Special Interest Group LISTSERV), searched the American Educational Research Association Online Repository, and cross-referenced using previous, published studies for conference abstracts or unpublished data. If the studies identified in the cross-referencing were not easily accessible, we contacted the primary authors to potentially obtain the study referenced. Of the studies that we retrieved in our search for grey literature, some studies had undergone some form of peer review (e.g., conference papers and posters, dissertations, theses) whereas others did not (e.g., unpublished manuscripts).

Several strings of search terms were used to obtain the articles included in the review. The scope of the terms ranged from general to theory-specific. A general search included the terms motivation and intervention and education, for instance. Other more specific searches included a motivation theory or construct in place of the term motivation. As an example, a search included expectancy-value theory...
and intervention and education. A multitude of iterations were also used such as motivation and only intervention or expectancy-value theory and only intervention. This same approach was used for each theory. Other search terms related to education included classroom, academic, learning, and school. In addition to direct searches, weekly email alerts from the search engines were created and sent to the first author. Both authors independently searched for studies to include in the meta-analysis. Specific search terms are listed in the Appendix found in the online version of the journal.

Procedures for Applying Inclusion Criteria

Both reviewers conducted the initial search in all sources and saved the results in an electronic format. During this stage, reviewers examined titles and abstracts. The initial search yielded 1,471 studies. The full texts for studies that met eligibility criteria were saved in an electronic file and those that obviously did not meet our search criteria were eliminated. For instance, some studies were strictly correlational, with no intervention delivered. If the relevance of the study was not clear on initial screening, then the reviewers retrieved the full texts and saved them in an electronic file. Both reviewers met periodically to discuss discrepancies. From the initial 1,471 studies, 158 studies were saved. The reviewers agreed on 153 for inclusion/exclusion, yielding a percent agreement of 96.8. Disagreements among the remaining five studies were resolved through discussion until a consensus was reached.

Selection Criteria and Sample of Studies

Inclusion Criteria

Papers that were read in full were coded using the following criteria: (a) the intervention must have been guided by a motivation theory (see Table 1); (b) there was an intervention that manipulated an independent variable, and the intervention included at least one type of naturalness according to Tunnell’s (1977) definition (i.e., natural treatment, natural setting, natural behavior, including studies with interventions delivered and data collected in the field setting as well as studies with interventions delivered in the lab but the dependent variables collected in the field); (c) the dependent variable represented an authentic educational outcome, including performance (e.g., standardized test scores), behaviors (e.g., persistence at a task, course choices), and self-reports (e.g., motivation, engagement); (d) the studies must have been an independent groups design with an intervention group and a comparison or control group. Dependent group, or repeated-measures, designs were not included in this meta-analysis because the appropriate effect size statistic (the standardized mean gain) is not comparable with the effect size statistic for independent group designs (the standardized mean difference; Lipsey & Wilson, 2001). The fifth criterion was the outcomes in the studies must have been collected on students spanning elementary through postsecondary schooling. Studies that only collected outcomes on teachers, but not students, were not included in the study.

Exclusion Criteria

Notably, three types of intervention studies based on motivation theory were excluded from our analyses. First, because of our focus on educational settings
<table>
<thead>
<tr>
<th>Theory/ framework</th>
<th>Description</th>
<th>Overview(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Achievement emotions</td>
<td>Emotional experiences in school emanate from students’ perception of control and value for academics</td>
<td>Pekrun (2006)</td>
</tr>
<tr>
<td>Anxiety</td>
<td>Worrying about the consequences of performance, which undermines working memory and outcomes</td>
<td>Ramirez and Beilock (2011)</td>
</tr>
<tr>
<td>Happiness</td>
<td>An overriding emotional sense of wellbeing</td>
<td>Fordyce (1977)</td>
</tr>
<tr>
<td>Achievement goal</td>
<td>Students’ goals for engaging in an activity shape how they approach, experience, and react to achievement situations</td>
<td>Elliot (2005); Kaplan and Maehr (2007)</td>
</tr>
<tr>
<td>Attribution</td>
<td>Students’ explanations for success or failure influence subsequent achievement behavior</td>
<td>Weiner (1980)</td>
</tr>
<tr>
<td>Expectancy-value</td>
<td>Student motivation is determined most proximally by success expectancies and perceived task value</td>
<td>Eccles et al. (1983)</td>
</tr>
<tr>
<td>Goal setting</td>
<td>Specific, difficult task goals produce higher commitment and performance than vague goals that are easy to attain</td>
<td>Locke and Latham (1990, 2002)</td>
</tr>
<tr>
<td>Implicit theories of intelligence</td>
<td>Students’ beliefs about whether intelligence is fixed (i.e., entity mindset) or is malleable (i.e., incremental mindset) influence goal striving, persistence, and performance</td>
<td>Dweck (1986, 1999)</td>
</tr>
<tr>
<td>Interest</td>
<td>The development and deepening of interest in specific topics and academics is influenced by situational and individual difference factors</td>
<td>Hidi and Renninger (2006)</td>
</tr>
<tr>
<td>Need for achievement</td>
<td>The importance of mastery, high achievement, and besting others to reach one’s full potential</td>
<td>McClelland, Atkinson, Clark, and Lowell (1976)</td>
</tr>
<tr>
<td>Possible selves</td>
<td>Students’ conception of what they might become (both desired and feared) serve as incentives for future behavior and a way to evaluate current behavior</td>
<td>Markus and Nurius (1986)</td>
</tr>
<tr>
<td>Self-affirmation</td>
<td>Students’ who perceive that they are in danger of confirming a stereotype about their group experience increased anxiety and reductions in performance</td>
<td>Steele (1988)</td>
</tr>
<tr>
<td>Self-confrontation</td>
<td>Students’ perception that their behaviors and values differ from their self-conception motivates change</td>
<td>Rokeach (1973)</td>
</tr>
</tbody>
</table>

(continued)
(including typical classroom settings, afterschool support programs, and school-wide programs), studies included in the review were specifically related to academic, educational, and/or motivational outcomes. Studies investigating topics such as physical activity, or drug and alcohol use, were excluded, although interventions in these areas are certainly abundant and available (e.g., Dudley, Okely, Pearson, & Cotton, 2012; Metcalf, Henley, & Wilkin, 2013). However, if the motivation interventions were delivered in the school setting and targeted health-related outcomes (e.g., stress), those studies were included in the meta-analysis (e.g., Vansteenkiste et al., 2005; Yeager, Johnson, et al., 2014).

Second, our definition of naturalness excluded some intervention studies from our review, even though they were focused on students in an education context. For example, Mueller and Dweck (1998) reported a series of six studies with elementary students that manipulated the feedback (i.e., praise) they received for doing an academic activity. Although these randomized interventions, based on Dweck’s theories of intelligence, positively affected student learning and motivation on the experimental activities, they were excluded from our review because children were pulled out of their normal classrooms to participate in the study (i.e., artificial setting) and the dependent variables were unique to the experiment and not a part of students’ normal classroom experience (i.e., artificial dependent variable).

Third, our review also excluded studies that targeted a broader range of social and emotional learning skills, such as First Things First (Connell et al., 2009) and Positive Action (Washburn et al., 2011). These studies were excluded because these interventions do not simply target motivational processes, but a host of other processes and school structures, such as administrative structures, curriculum, and pedagogy (Rimm-Kaufman & Hulleman, 2015). As such, it is difficult to disentangle the effects of the motivational aspects of these broader initiatives from other

<table>
<thead>
<tr>
<th>Theory/framework</th>
<th>Description</th>
<th>Overview(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-determination</td>
<td>Satisfying students’ three core needs (autonomy, relatedness, competence) are essential for promoting motivation and well-being</td>
<td>Deci and Ryan (1985)</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>Students’ perception that they can successfully complete the specific tasks and activities required for learning promotes learning outcomes</td>
<td>Bandura (1982)</td>
</tr>
<tr>
<td>Social belongingness</td>
<td>The degree to which students perceive they belong and are connected to others can influence their learning outcomes</td>
<td>Baumeister and Leary (1995)</td>
</tr>
<tr>
<td>Transformative experience</td>
<td>Reframing the learning experience as an application of the content in a way that enhances everyday value</td>
<td>Pugh (2011)</td>
</tr>
</tbody>
</table>
elements, such as study strategies, emotional awareness, and decision making (for a review, see Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011).

Coding of Study Characteristics Used for Moderator Analyses

Both authors were involved in the development of the data extraction and coding form that was created in Microsoft Excel. This form also included a key for the various codes, and where necessary, an operational definition of the code to better ensure mutual understanding. For instance, we operationally defined elementary school as Grades K-5, middle school as 6 to 8, high school as 9 to 12, and postsecondary as community college, undergraduate university/college, and graduate school, unless otherwise noted by the authors. For example, if a sample included students in 6th grade but the authors explicitly noted the setting as elementary school, this sample was coded as elementary aged (e.g., Craven, Marsh, & Debus, 1991).

We pilot tested the coding form using a sample of included studies and discussed any codes that were still unclear. Following the initial pilot testing, both authors independently coded 100% of the included studies, and we compared our coding to identify and discuss discrepancies. Disagreements were resolved through consensus. Interrater reliability was assessed via percent agreement and reported in the moderator descriptions below.

Coding of Moderators

One purpose of a meta-analysis is to examine the moderation of effect sizes due to study characteristics. Because of the limited number of intervention studies in our meta-analysis, the moderator analysis is necessarily limited by several factors, including statistical power and likely confounding of our measured and unmeasured moderators. With this in mind, we considered five types of moderator variables in our review: theoretical framework of the intervention, grade level of participants, type of dependent measure, experimental design, and degree of naturalness.

Theoretical Framework

The theoretical framework that undergirds the development of the intervention is important to consider because theories differ in terms of which underlying psychological mechanism they target. As outlined in Table 2, there were 15 different theoretical frameworks of motivation represented by the interventions in our review. The most prevalent theoretical perspectives that guided the motivation interventions in our review were attribution theory (13 studies; e.g., Wilson & Linville, 1982), self-determination theory (11 studies; e.g., Reeve et al., 2004), and self-affirmation theory (8 studies; e.g., Cohen et al., 2006). Other theoretical frameworks with multiple studies included achievement emotions (7 studies; e.g., Fordyce, 1977), expectancy-value (7 studies; e.g., Hulleman & Harackiewicz, 2009), implicit theories of intelligence (6 studies; e.g., Blackwell et al., 2007), social belongingness (5 studies; e.g., Walton & Cohen, 2007), achievement goals (4 studies; e.g., Muis et al., 2013), transformative experiences (4 studies; e.g., Pugh, 2002), possible selves (3 studies; e.g., Oyserman et al., 2006), interest theory (2 studies; e.g., Guthrie et al., 2006), and several theories with one study (goal...
setting, need for achievement, and self-confrontation). If two or more theories contributed to the development of the intervention, then the study was coded as multiple perspectives, which was the most common type of study in our review (23 studies). Interrater agreement for coding of theoretical framework was 87%.

**Student Grade Level**

Responsiveness to interventions may have a developmental component. Younger or older students may respond to some interventions to a greater or lesser extent than others (Wentzel & Wigfield, 2007). Ideally, this hypothesis would be tested through applying the same intervention across multiple grade levels. In our sample of studies, only interventions grounded in self-determination theory have been tested in samples ranging from elementary school to postsecondary aged students. Most other interventions were tested in one or two age ranges (e.g., social belongingness interventions have been examined only in college students and expectancy-value interventions have been tested in high school and college samples). Table 3 includes the student grade levels coded for each study. Studies
TABLE 3
Number of studies (k), mean effect sizes (average d), and confidence intervals (95% CI) for moderator analyses

<table>
<thead>
<tr>
<th>Moderator type</th>
<th>k</th>
<th>Average d</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student grade level</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary school (K–Grade 5)</td>
<td>10</td>
<td>0.52</td>
<td>[0.31, 0.73]</td>
</tr>
<tr>
<td>Middle school (Grades 6–8)</td>
<td>24</td>
<td>0.57</td>
<td>[0.44, 0.69]</td>
</tr>
<tr>
<td>High school (Grades 9–12)</td>
<td>17</td>
<td>0.42</td>
<td>[0.27, 0.57]</td>
</tr>
<tr>
<td>Postsecondary</td>
<td>41</td>
<td>0.47</td>
<td>[0.38, 0.57]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>92</td>
<td>0.49</td>
<td>[0.43, 0.56]</td>
</tr>
<tr>
<td><strong>Type of experimental design</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized</td>
<td>64</td>
<td>0.43</td>
<td>[0.36, 0.50]</td>
</tr>
<tr>
<td>Quasi-experimental</td>
<td>28</td>
<td>0.64</td>
<td>[0.52, 0.75]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>92</td>
<td>0.49</td>
<td>[0.43, 0.56]</td>
</tr>
<tr>
<td><strong>Degree of naturalness—Dependent variable (DV)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DV in educational context</td>
<td>70</td>
<td>0.46</td>
<td>[0.39, 0.53]</td>
</tr>
<tr>
<td>DV not in educational context</td>
<td>22</td>
<td>0.63</td>
<td>[0.48, 0.77]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>92</td>
<td>0.49</td>
<td>[0.43, 0.56]</td>
</tr>
<tr>
<td><strong>Degree of naturalness—One vs. two vs. three degrees</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One degree present</td>
<td>7</td>
<td>0.56</td>
<td>[0.30, 0.82]</td>
</tr>
<tr>
<td>Two degrees present</td>
<td>28</td>
<td>0.46</td>
<td>[0.34, 0.59]</td>
</tr>
<tr>
<td>Three degrees present</td>
<td>57</td>
<td>0.50</td>
<td>[0.42, 0.58]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>92</td>
<td>0.49</td>
<td>[0.42, 0.56]</td>
</tr>
<tr>
<td><strong>Type of dependent variable</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-report</td>
<td>75</td>
<td>0.54</td>
<td>[0.46, 0.63]</td>
</tr>
<tr>
<td>Performance indicator</td>
<td>61</td>
<td>0.52</td>
<td>[0.43, 0.51]</td>
</tr>
<tr>
<td>Behavioral indicator</td>
<td>28</td>
<td>0.62</td>
<td>[0.49, 0.76]</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>164</td>
<td>0.55*</td>
<td>[0.47, 0.58]</td>
</tr>
</tbody>
</table>

*The overall weighted mean effect size for type of dependent variable is slightly different than the other moderators because multiple effect sizes from the same study were analyzed as unique effect sizes (e.g., a study with both self-report and performance outcomes were analyzed separately). Alternatively, for the other moderators, the average of multiple effect sizes within each study was used.

were coded according to the grade level participants were in during the study and included the following: elementary school (K to Grade 5), middle school (Grades 6–8), high school (Grades 9–12), and postsecondary (community college, undergraduate university/college, graduate school). Interrater agreement for coding of student grade level was 98%.

**Type of Dependent Variable**

In our review of studies, we found three basic types of dependent measures: participant self-report (e.g., interest, achievement goals), performance outcomes (e.g., standardized test scores, course grades), and behavioral outcomes (e.g., discipline referrals, retention). It is possible that some dependent measures will be
easier to affect than others. For example, to the extent that behavioral and performance measures are multiply determined (Duncan et al., 2007; Grissmer, Grimm, Aiyer, Murrah, & Steele, 2010), then these measures might be more difficult to influence than self-report measures. In addition, to the extent that self-report measures are more closely aligned with the theorized mechanism, then intervention impacts might be greater on self-report measures than on behavioral and performance outcomes.

Table 3 includes the type of dependent variables coded for each study. When coding behavioral outcomes, we changed the direction of the effect size such that positive values indicated an increase in preferred behaviors. For instance, if the treatment group had a larger decrease in discipline referrals following the intervention compared with the control group, the direction of this effect was changed from negative to positive. If a single study included multiple types of dependent variables, each type of dependent variable was coded separately, yielding an effect size average for each type. Because of this manner of coding the effect sizes for this moderator, there were 164 effect sizes, compared with 92 for the other moderator analyses. Interrater agreement for coding the type of dependent variables was 97%.

**Experimental Design**

In order to be included in our review, studies had to have some type of control or comparison group. Studies that randomly assigned participants to treatment and control conditions were coded as experimental, whereas studies that did not use random assignment were coded as quasi-experimental. Table 3 includes the type of experimental design coded for each study. Interrater agreement for coding of experimental design was 93%.

**Degree of Naturalness**

The extent to which an intervention is tested within an ecological context is a potentially important moderator. For example, interventions that are delivered in the laboratory are likely to have a higher degree of implementation quality than those delivered in the field, and thus produce larger effect sizes (Hulleman & Cordray, 2009). Based on Tunnell’s (1977) definitions, we coded studies based on whether the intervention was part of the regular academic experience (natural treatment), occurred in a setting outside the laboratory (natural setting), or included a dependent measure that normally occurs within the educational context (e.g., exams, choices about activities; natural behavior). Table 3 includes the degree of naturalness coded for each study. Each study had to include at least one type of naturalness to be included in our analyses. Interrater agreement for coding the degree of naturalness was 86%.

**Coding of Effect Sizes**

A final sample based on our inclusion criteria yielded 74 usable papers which contributed 92 independent studies (see Supplementary Table S1 for a narrative description of the studies and Supplementary Table S2 for effect sizes and study characteristics, available in the online version of the journal). If one paper contained multiple studies, each study was coded separately. For instance,
Supplementary Table S2 shows the Wilson and Linville (1985) paper that included two separate replications of their original 1982 study (i.e., distinct samples were used for each study), and therefore, we calculated the average effect size for each study. If a single study included multiple effect sizes (e.g., the Acee & Weinstein [2010] study reported intervention effects on task value, endogenous instrumentality, choice-behavior, and exam performance), then those effect sizes were averaged. Although averaging those effect sizes in our analyses violate the assumption of independence, Baldwin and Shadish (2011) demonstrated that computing the simple average effect sizes as we did yields similar results compared with approaches that account for dependency in measures.

Averaging the simple effect sizes resulted in 92 independent effect sizes for the moderator analyses on theoretical framework, student grade level, experimental design, and degree of naturalness. The exception was for the moderator analysis on type of dependent variable. If one study had three different types of dependent variables, then that one study contributed three separate variables and associated effect sizes. Coding the studies in this manner resulted in 164 effect sizes for the type of dependent variable moderator analysis. Because we analyzed our moderators using separate ANOVAs, we adjusted our critical $p$ value to a more conservative value using the Bonferroni correction. This resulted in a critical $p$ value of .01, rather than .05.

**Calculation of Effect Sizes**

Different studies in our meta-analysis commonly used different instruments to measure a dependent variable or construct of interest and thus were not numerically comparable across studies. This included situations where the same construct was operationalized in a different manner or when different constructs were measured across studies. In both these instances, the effect size statistic used to aggregate the findings in the meta-analysis was standardized so that the values on the original measures were comparable. To do so, we used Cohen’s $d$ as the effect size statistic in this meta-analysis, which is the standardized difference between two means (Cohen, 1988) and reflects the degree to which the difference is practically significant or meaningful. Cohen’s $d$ is commonly used to compare an experimental group with a control group on one or more continuous dependent variables. We calculated this effect size using the formula recommended by Lipsey and Wilson (2001),

$$ES = \frac{\bar{X}_{G1} - \bar{X}_{G2}}{S_{pooled}},$$

where $\bar{X}_{G1}$ and $\bar{X}_{G2}$ are the means for Groups 1 and 2, respectively, and $S_{pooled}$ reflects the pooled standard deviation (Hedges & Olkin, 1985). If researchers did not directly report values for Cohen’s $d$ in a study, we calculated these values with the reported means and standard deviations. If sufficient information was not available (e.g., standard deviations were not reported), we attempted to contact the authors directly for this information. This option was only necessary for three studies. We also calculated $d$ if certain statistics required a conversion (e.g., $t$ statistic to $d$). For convenience, we interpreted the magnitude of Cohen’s $d$
according to the conventions recommended by Cohen (1988): \(\leq 0.20\) was considered small, \(0.50\) was considered medium, and \(\geq 0.80\) was considered large. For example, a Cohen’s \(d\) of \(0.50\) suggests that the difference between the experimental and control conditions was one half of a standard deviation on the outcome measure. This effect size would be considered a medium effect.

### Results

#### Overview

The data were analyzed in three ways. First, the effect sizes were submitted to a meta-analysis using procedures outlined by Lipsey and Wilson (2001) and explained below. We investigated the moderator variables of theoretical framework, sample characteristics, type of dependent variable, experimental design, and degree of naturalness. Mean effect size and moderator analyses were conducted via macros provided by Lipsey and Wilson (2001) using IBM SPSS Statistics for Windows, Version 21 (IBM Corp., 2012). Second, we conducted a sensitivity analysis by identifying outliers and re-running the analyses with those effect sizes removed. Third, although we minimized the likelihood of publication bias by including grey literature, we investigated potential publication bias by examining a funnel plot (see Supplementary Figure S1 available in the online version of the journal), calculating Orwin’s (1983) fail-safe \(N\), and conducting a moderator analysis comparing published studies and unpublished studies.

Because we were interested in making inferences to the population of studies from which our studies are considered to be a random sample, a random effects model was utilized in calculating the overall weighted effect size. Random effects models differ from fixed effects models in a number of ways. The random effects model allows the true effect size to vary depending on the study, whereas the fixed effects model assumes that all studies comprising the meta-analysis share one common effect size (Hedges & Vevea, 1998). In a fixed effects model, fluctuations in effect sizes are presumed to be due to within study estimation error (sampling variance) alone. However, in a random effects model, sources of variation include both the within study estimation error and between studies variance (Borenstein, Hedges, Higgins, & Rothstein, 2010). Differences between the random effects and fixed effects models also have implications regarding the inferences that can be drawn from the results. A random effects model allows the researcher to generalize results beyond those found in the study, whereas this generalization is inappropriate for a fixed effects model. For a fixed effects model, one must have strong evidence that the studies included in the meta-analysis were virtually identical (Aronson, Ellsworth, Carlsmith, & Gonzalez, 1990).

Schmidt, Oh, and Hayes (2009) provided an example, noting,

If the studies drew their samples from the same population (e.g., college sophomores), tested exactly the same hypotheses with exactly the same study design, treatment strength (if an experimental study), measures, instructions, time limits, etc., then one might assume \emph{a priori} that the same population parameter was estimated in all the primary studies (i.e., \(\sigma_\delta^2 = 0\) or \(\sigma_\rho^2 = 0\)) and this could be the basis for choosing the [fixed effects] model. (p. 124)
This assumption is not tenable in the current meta-analysis. Therefore, we chose to analyze our data using the random effects model because (a) we could not meet this strict assumption in the fixed effects model, (b) we hypothesized the true effect size would vary from study to study, and (c) we wanted to appropriately draw inferences to studies beyond those included in our meta-analysis.

**Outliers and Effect Size Adjustments**

Prior to running any analyses, we examined the distribution of effect sizes to determine whether outliers were present in the meta-analysis. A scatter plot of the effect sizes across studies was visually inspected. Four outliers were identified that were considerably larger than the sample of studies as a whole. Rather than simply eliminating these outliers, we recoded the effect sizes to more moderate ones using the Windsorizing procedure described by Lipsey and Wilson (2001). To do this, each effect size that was greater than three standard deviations from the mean effect size was recoded to the value at three standard deviations. The overall mean effect sizes were re-run with these four Windsorized effect sizes to determine the influence of these outliers on our results. The overall weighted mean effect size decreased only slightly with these adjusted effect sizes ($d = 0.49$ compared with 0.51 in the original analyses). All subsequent analyses were conducted using these adjusted effect sizes.

**Meta-Analysis**

The overall weighted mean effect size using a random effects model was 0.49 (95% CI [0.43, 0.56]; $z = 14.69$, $p < .001$). Aside from 12 studies with effect sizes either negative or small, effect sizes for most motivation interventions ranged from moderate ($d = 0.27$ to 0.58) to large ($d = 0.61$ to 1.28), which is considerably high for interventions related to education (Borman, Hewes, Overman, & Brown, 2003).

To assess homogeneity in the sample of effect sizes, we computed the random effects variance component, $v_\theta$, which represents the degree of heterogeneity in effect sizes (the variance among the true effect sizes in the population). The associated homogeneity statistic, $Q$, indicates whether $v_\theta$ is significantly different than 0, and if significant, then this indicates heterogeneity or significant variation among the effect sizes in the population. The overall homogeneity statistic, $Q$, was statistically significant, $\chi^2(91) = 451.06$, $p < .001$. Thus, we rejected the hypothesis of homogeneity and concluded that the variance in the population of effect sizes was greater than would be expected from sampling error alone. In addition to $Q$, we also computed $I^2$, which reflects the proportion of the variation in effect size estimates that is due to heterogeneity rather than by chance (Higgins & Thompson, 2002). The formula for this statistic is represented in Equation 2 below, where $k$ represents the number of studies and $Q$ is the homogeneity statistic:

$$ I^2 = 100\% \times \left( \frac{Q - (k - 1)}{Q} \right) $$

Higgins and Thompson (2002) recommended the following interpretative guidelines for $I^2$: 25% (small heterogeneity), 50% (medium heterogeneity), and
75% (large heterogeneity). For the overall homogeneity statistic above ($Q = 451.06$) with $(k - 1)$ studies (91), $I^2 = 80\%$, which reflects a large degree of heterogeneity.

**Moderator Analyses**

To account for excess variability in effect sizes, we used the Inverse Variance Weighted one-way ANOVA (random effects model estimated via iterative maximum likelihood) for our moderator analyses (Lipsey & Wilson, 2001). In this approach, each effect size is treated as an observation, and variance in effect sizes across studies is analyzed in the ANOVA framework. For the following analyses, the overall $Q$ statistic was partitioned into two types of variance, $Q_B$ and $Q_W$. $Q_B$ represents the between-group variance in effect size, and $Q_W$ represents the within-group variance in effect size. In the moderator analyses, if $Q_B$ was significant, this indicated that there were significant differences in effect sizes across groups. If $Q_B$ was not significant, then the moderator did not account for a significant amount of the variability in the effect sizes. When $Q_B$ was significant, we also examined $Q_W$. If $Q_W$ was significant, this indicated that there was additional variance in effect sizes not explained by the moderator. However, if $Q_W$ was not significant, the moderator was considered adequate in accounting for excess variability in the effect size distribution (Lipsey & Wilson, 2001).

Tables 2 and 3 show the mean effect sizes, their 95% confidence intervals, and number of studies for each moderator.

**Theoretical framework.** Average effect sizes for theoretical framework were largest for transformative experiences ($d = 0.74$) and smallest for social belongingness ($d = 0.35$). However, the $Q_B$ statistic was not significant, $\chi^2(11) = 16.82, p = .1133$, indicating that studies’ effect sizes did not differ based on theoretical framework.

**Student grade level.** Average effect sizes for student grade level were largest for students in middle school (Grades 6–8, $d = 0.57$), followed by students in elementary school (Grades 1–5, $d = 0.52$), postsecondary institutions ($d = 0.47$), and high school (Grades 9–12, $d = 0.42$). However, the $Q_B$ statistic was not significant, $\chi^2(3) = 2.46, p = .4823$, indicating that studies’ effect sizes did not differ based on student grade level.

**Type of dependent variable.** Average effect sizes for type of dependent variable were larger for behavioral ($d = 0.62$) than self-reports ($d = 0.54$) or performance indicators ($d = 0.52$). However, the $Q_B$ was not significant, $\chi^2(2) = 1.59, p = .4527$, indicating that studies’ effect sizes did not differ based on the type of dependent variable measured.

**Experimental design.** Average effect sizes for experimental design were larger for quasi-experimental designs ($d = 0.64$) compared with randomized experimental designs ($d = 0.43$). The $Q_B$ statistic was significant, $\chi^2(1) = 9.06, p < .01$, indicating
that studies’ effect sizes differed based on the type of experimental design. Furthermore, the pooled within group variance was homogeneous ($Q_w = 86.74, p = .5776$), suggesting the categorical variable represented in $Q_B$ (experimental design) accounted for the excess variability in effect sizes.

**Degree of naturalness.** Two subanalyses were conducted for degree of naturalness: whether a study included a dependent variable that normally occurs within the educational context, and whether a study included one, two, or all three degrees of naturalness. The $Q_B$ statistic was not significant for either subanalysis, $\chi^2(1) = 3.95, p = .0467; \chi^2(2) = 0.58, p = .7473$, indicating that effect sizes did not differ based on degree of naturalness in the study.

**Publication Bias**

Although the inclusion of grey literature can diminish the likelihood of publication bias, we also examined this possibility through three statistical methods. The first method, the funnel plot (Light & Pillemer, 1984), is the most commonly used method to detect publication bias (Torgerson, 2006). Funnel plots graphically depict a point estimate for each study on the $x$-axis (usually the effect size) against a measure of the precision for each study on the $y$-axis (usually the sample size or standard error). When little to no publication bias is present, the data points on the graph will look like an inverted funnel. However, when publication bias may be present, the left side of the funnel will have missing data points, which would depict the absence of negative or null results among the studies. As presented in Supplemental Figure S1 (available in the online version of the journal), the funnel plot shows that studies with smaller sample sizes are distributed around the mean effect size on both sides of the distribution, thus suggesting minimal publication bias.

The second method is the fail-safe $n$ test, which was first developed by Rosenthal (1979) for $z$ values, then adapted by Orwin (1983) for use with the standardized mean difference effect size. Orwin’s fail-safe $n$ approach estimates “the number of studies with an effect size of zero needed to reduce the mean effect size to a specified or criterion level” (Lipsey & Wilson, 2001, p. 166). To calculate the fail-safe $n$, the researcher must determine a criterion effect size that would be too small to be of theoretical or practical significance. Orwin recommended a $d$ of 0.20 as the criterion effect size, which reflects the magnitude of an effect size conventionally considered to be small. The fail-safe $n$ is calculated using the following formula,

\[
N_{fs} = \frac{N(d - d_c)}{d_c}
\]

where $N$ represents the number of studies in the meta-analysis, $d$ represents the average effect size for the studies in the meta-analysis, and $d_c$ represents the criterion value selected that $d$ would equal (typically 0.20) when the number of hypothetical studies ($N_{fs}$) were added to the meta-analysis. Therefore, $N_{fs}$ equals the number of hypothetical studies necessary to change the obtained effect size ($d$) into a small effect size with little to no theoretical or practical significance. With
our meta-analysis of 92 published studies and weighted effect size of 0.49, Orwin’s fail-safe \( n \) suggests that an additional 133 studies with a mean effect size of zero would be needed to reduce the mean effect size to 0.20.

Third, we conducted a moderator analysis comparing published to unpublished studies using the same Inverse Variance Weighted one-way ANOVA used for our other moderator analyses. As expected and congruent with previous research (Rothstein & Hopewell, 2009), effect sizes were larger for published studies (\( d = 0.54 \)) compared with unpublished studies (\( d = 0.26 \)). The \( Q_B \) statistic was significant, \( \chi^2(1) = 11.00, p < .001 \), indicating that effect sizes differed based on whether the studies were published or unpublished. Furthermore, the pooled within group variance was homogeneous (\( Q_W = 84.30, p = .6497 \)), suggesting the categorical variable represented in \( Q_B \) (publication status) was sufficient to account for the excess variability in the effect size distribution for this analysis. This finding underscores our decision to search for and include grey literature in this study. Had we only included published articles, we would have overestimated the true effect size; thus, the addition of unpublished studies in this meta-analysis likely yields a more accurate estimate of the true effect of the motivation interventions in education. Furthermore, excluding grey literature would have attenuated the breadth of coverage of the available evidence and thus may have introduced systematic error and posed a threat to the validity of our findings (Moher, Cook, Eastwood, Olkin, & Stroop, 2000).

**Discussion**

Results from this meta-analytic review indicate that motivation interventions have demonstrated promising results for enhancing educational outcomes. We meta-analyzed 74 published and unpublished papers of 92 field studies grounded in motivation theory, accounting for 38,377 participants. The motivation interventions in this review were effective overall, averaging approximately a half a standard deviation effect size (\( d = 0.49 \)). Importantly, this average effect size did not significantly vary according to any of our moderator analyses except for experimental design, with randomized experiments demonstrating smaller effect sizes than quasi-experiments. In other words, given the studies in our sample, there were no statistically significant differences in effect size due to theoretical framework of the intervention, age of participants (elementary through postsecondary students), type of dependent variable (performance, behavior, self-reported motivation), or degree of naturalness. In addition, the studies included in our review spanned a range of sample characteristics associated with diversity, including ethnicity, culture, socioeconomic status, and ability/disability. The samples also included elementary- to postsecondary-aged students. Thus, the review documents relatively robust effects of motivation interventions across diverse samples.

**Implications for Theory Development**

Although the moderator analysis by theoretical framework was not statistically significant, an inspection of Table 2 demonstrates variability in intervention effect sizes based on theory. Should we conclude from our review that some theories produce stronger effects than others? We would urge caution when making such an interpretation. First, and perhaps most statistically substantive, the confidence
intervals among the effect sizes in our comparisons were quite large. For instance, in Table 2, interest theory with only two studies, has a confidence interval spanning 0.30 (small effect) to 1.08 (large effect). These large confidence intervals reflected less precision, decreased power, and may have affected our ability to find a significant result when in fact one might have existed (i.e., Type II error). Second, not only was the homogeneity statistic not significant, but the measure of effect size was small. That is, the moderator analysis was not statistically or practically significant. Third, the delivery mechanisms of the interventions varied widely across theory and population, which may have resulted in differences in intervention strength due to differences in dosage, frequency, and timing (O’Donnell, 2008). For example, several interventions were imbedded within the context of week-long curricula (e.g., Feng & Tuan, 2005; Pugh, 2002), whereas other interventions were independent activities requiring as little as 20 minutes for students to complete (e.g., Cohen et al., 2006; Hulleman & Harackiewicz, 2009).

If we cannot make inferences about relative importance and potency among theories and intervention, then what can we conclude? First, educational psychologists and other education researchers should be emboldened by the fact that their ideas, played out through interventions, can have meaningful change on educational outcomes. Overall, the theories that inspired these interventions and were tested experimentally have effected positive change in students’ educational outcomes. Second, the theoretical models tested in our review should carry additional weight when considering which psychological constructs and processes are most important in education. Not only do these models have correlational and observational support, but the interventions designed to target specific psychological processes affected outcomes.

Third, by focusing on motivation interventions, we learned that motivation can be a key process or mechanism for enhancing student learning outcomes. Certainly, helping students develop optimal motivation is an important goal in its own right; in fact, many argue that it is even more important than standardized metrics such as test scores (Eccles et al., 1983; Hidi & Harackiewicz, 2000; Nicholls, 1979; Ryan & Brown, 2005). However, motivation is also a key process that enables learning (Hidi, 1990). Our review demonstrates that these constructs and their processes deserve additional attention, both conceptually and empirically. For example, effective interventions could be combined so that their additive and synergistic effects could be tested, either as supercharged interventions or as part of deeper curricular reforms. Such work requires more than simply putting two interventions together, such as sequencing an anxiety intervention prior to a value intervention. Instead, careful conceptual work is required to understand which interventions and constructs amplify each other, and how they connect developmentally (i.e., does one come before the other).

Fourth, the studies found in this review provide a strong foundation for understanding causal relationships between motivation constructs and educational outcomes. Many of the correlational studies excluded from our review drew causal conclusions based on nonexperimental data, a problematic finding echoed by Robinson et al. (2007) in their review of published studies in the area of teaching and learning. Experimentally testing theories through intervention studies provides...
further validity evidence for the theory. For example, randomized experiments support causal inference in a way that many other research methods cannot.

Fifth, as measurement of motivational constructs and data analytic techniques continue to improve, those interested in intervention research should capitalize on these improvements to provide additional validity information for motivation theories. Furthermore, the results substantiate the potential power of motivation interventions on educational outcomes beyond self-report surveys or instruments. Our results indicated no differences among self-report, performance, or behavioral outcomes. This is particularly important given the policy-relevant nature of behavioral and performance outcomes (e.g., achievement gaps, retaining STEM majors), and provides a crucial link between theory and policy. As a result of this work, findings will not only become more generalizable and yield more meaningful recommendations for practice, but also be based on broader scientific evidence.

Finally, our review of intervention studies targeting student motivation provides evidence of how practice can be improved. Rather than providing hypotheses about what should work, intervention studies provide evidence of what can work (given the particular circumstances of the study). By strengthening the reciprocal relationship between theory, research, and practice, field interventions bolster practical validity arguments integrating psychological theories into the mainstream of educational practice (Hulleman & Barron, 2016). Correlational studies or lab studies certainly inform one another as theory guides research and results of that research then informs theory. In contrast, field intervention studies can also provide valuable information for theory development as well. If theories are tested more extensively in the field or in practice, results can then guide theoretical developments and improvements, which then influence subsequent research and practice in a meaningful way (Walton, 2014).

**Implications for Policy, Practice, and Research**

**Policy and Practice Implications**

Findings from this meta-analysis have clear relevance for state and local policies concerning approaches to enhance student learning and achievement. Policy makers, researchers, and educators are focused on accountability and ways to increase student performance and persistence in school, and to better prepare students for future education and entry into an increasingly global workforce (Silva & White, 2013). A well-educated workforce is critical for maintaining our country’s economic competitiveness (Global Science Forum, 2006). Hence, there is a need for data gathered through rigorous methods to accomplish this objective (Easton, 2013), and that informs educators who desire to improve student success. By targeting psychological mechanisms, we can enhance educational outcomes in an efficient and cost-effective manner. In contrast to comprehensive school reform models, which can cost thousands of dollars per student per year to implement (Borman et al., 2003), these motivation interventions are extremely cost effective (Yeager & Walton, 2011). In fact, most of the motivation interventions in our review cost little or no money to implement (e.g., the self-affirmation intervention and utility value interventions cost nothing more than the time to deliver them to students via computer or paper-and-pencil), and can be implemented across content areas and contexts. Notably, the average effect size on performance indicators
in our review was 0.52, whereas the average effect size of comprehensive school reform models was 0.11 (Borman et al., 2003).

There is an important caveat to this optimism. Translating the interventions tested here into educational practices is not always straightforward. Teachers often lack the research expertise necessary to adapt emerging principles from the research literature, and researchers often underestimate the complexity of classrooms because they are not fully embedded within school contexts. Thus, a researcher-designed intervention found to be effective in the lab or in a specific field context may not work when applied to another context (Daniel, 2012). This concern may be because teachers were unable to deliver the intervention with fidelity, or because researchers did not account for additional moderating variables of a particular educational context or subpopulation of students (O’Donnell, 2008). Some interventions are designed with this translation in mind, such as Reeve’s teacher professional development program. However, other interventions are designed to be independent of teaching practices, such as attribution retraining programs or the self-affirmation intervention. It is less clear how teachers should incorporate these interventions to enhance their teaching practices. Should we design professional development for teachers focused on helping teachers imbed these interventions within their teaching? Should school-wide reform initiatives be created based on these constructs and interventions? Or, would it be more effective to staff schools with psychological engineers who, armed with the knowledge and skills to implement an array of interventions, could be responsible for affecting student motivation (Yeager & Walton, 2011)?

Regardless of the solution, it is unlikely that teachers working alone or researchers working alone will be able to effectively and efficiently translate research results into scalable classroom practices. What would this work look like? As one example, the Carnegie Foundation for the Advancement of Teaching has created several researcher–practitioner collaborations that utilize the tools of improvement science to address important practical problems (e.g., Berwick, 2008; Bryk et al., 2015). Researchers and practitioners are equal participants in this model to effectively address educational challenges that neither group could solve alone. For example, in the Student Agency Improvement Community, researchers’ background expertise in social–psychological theory and the development and testing of interventions are joined with the wisdom and experience of teachers and administrators (Hulleman & Barron, 2016). The goal is to develop effective and feasible classroom-based curricular innovations based on effective social psychological interventions. Other collaborations include addressing the high failure rate of community college remedial mathematics students (Silva & White, 2013), and the high attrition rate of early career teachers (Headden, 2014).

Research Implications

Although the results from this meta-analysis can begin to shed some light on the answers to these types of questions, as a field we need to establish mechanisms and methods for doing this translational research. As researchers, we receive very little training (if any!) on how to use our results to inform policy makers and practitioners about promising constructs, interventions, and initiatives that can improve motivation and educational performance in our schools. If our work is to have an impact on practice, not only do we need to conduct research
that produces appropriate scientific evidence (Mortensen & Cialdini, 2010), but we also need methods and processes of translation.

Furthermore, we need more intervention research, design-based and experimental, that develops and tests operationalized theories in education contexts. It is only through more and continued energy in this area will be make gains in both our theoretical models and practical ramifications. Given that 45% of the intervention studies used postsecondary-aged samples, we need more K-12 interventions. Researchers need to examine whether early interventions might be more or less powerful, or whether there is a sensitive period for optimal intervention effectiveness.

Limitations and Future Directions

One of the limitations of the current study was the dearth of motivation intervention field studies in education. Despite the effectiveness of these interventions, far fewer intervention field studies grounded in motivation theory have been conducted relative to correlational/nonexperimental studies and laboratory studies. Given the extensive body of research and theorizing on motivation in education, there were a disappointingly small amount of field studies using motivation interventions. This limitation tempers the strength of the inferences we can draw from our data, especially regarding the moderator analyses. For instance, despite apparent differences in effect sizes based on theoretical framework, neither statistical nor practical differences among theories were found. This finding may be because some theories contained two or three studies in this meta-analysis and therefore our tests lacked sufficient statistical power. In addition, theoretical frameworks were not equally distributed across grades, so it is possible that grade level and theoretical framework are confounded. It is plausible that other variables within a particular set of studies for a particular theory may have affected our results. As an example, if a group of studies for one theory all had higher levels of intervention dosage (e.g., an intervention that comprised of 8 total hours, over several weeks), one might assume the effect would be larger compared with a group of studies for another theory with lower levels of intervention dosage (e.g., a 10-minute writing activity). Coding the studies to account for intervention dosage will be an important next step for future research.

Future meta-analyses could also include laboratory studies of motivation interventions. Doing so would not only help increase the sample size overall, but also in comparing the relative effectiveness of field versus lab studies (Hulleman & Cordray, 2009). It is often assumed that laboratory studies will have stronger effects than their counterparts in the field. However, little empirical evidence exists on the diffusion of such transfer effects. The drawback to including laboratory studies is that the transfer of effects from the laboratory to the field is often a dubious endeavor based on contextual differences (Daniel, 2012), or to changes necessitated by new settings, students, and implementers (Datnow & Stringfield, 2000).

Conclusion

Across a diverse range of sample characteristics, including ethnicity, culture, socioeconomic status, grade level, and ability/disability, this meta-analytic review documents the effectiveness of educational interventions designed to enhance student motivation. We hope this review encourages more researchers to engage in intervention research. When combined with other methods, intervention research
provides the opportunity for researchers to create a set of recommendations for practices based on appropriate scientific evidence. Without intervention research, educational researchers are needlessly left on the sidelines when educators ask, “What should I do now in my classroom based on your research?”

Notes

Rory A. Lazowski is now an assistant research scientist at The College Board.

This research was supported by the National Science Foundation, through Grant DRL 1252463 to the second author.

The authors would like to thank Kenn Barron, Jamie DeCoster, Hunter Gehlbach, and Dena Pastor for their helpful comments on prior versions of this article.

1 We use the term *appropriate scientific evidence* to highlight that a range of evidence is required for us to make claims about what practitioners should do. The effect of an intervention is a theoretically different relation than the baseline correlation between two measured variables. Therefore, it is not acceptable to conduct an observational study and then make broad recommendations for practice without first conducting appropriate intervention studies.

2 Although quasi-experiments and experiments are not the only source of evidence for causality (longitudinal, qualitative, and mixed-methods research are also crucial), experimental methods are an important and underexamined aspect of building the case for internal validity in the field of motivation. Without internal validity evidence in support of our recommendations, we have failed to establish a base of appropriate scientific evidence to make such recommendations.

References

*References marked with an asterisk indicate studies included in the meta-analysis.*


Motivation Interventions

*Educational Psychology: An International Journal of Experimental Educational Psychology. Advance online publication. doi:10.1080/01443410.2014.895293


References


Motivation Interventions


Lazowski & Hulleman


Motivation Interventions


Motivation Interventions


Motivation Interventions


Authors

RORY A. LAZOWSKI is an assistant research scientist at The College Board, 661 Penn Street, Newtown, PA 18940; e-mail: rlazowski@collegeboard.org. He was previously an instructional faculty member and coordinator for the Peer Assisted Study Sessions in the Learning Centers at James Madison University. He recently received his PhD in assessment & measurement at James Madison University and has research interests in the areas of academic motivation and meta-analysis.

CHRIS S. HULLEMAN is a research associate professor of applied developmental science in the Curry School of Education at the University of Virginia, 405 Emmet St S, Charlottesville, VA 22904; e-mail: chris.hulleman@virginia.edu. He is a fellow of the Carnegie Foundation for the Advancement of Teaching. His work focuses on psychological theories and interventions to increase student motivation and development of a framework to assess intervention fidelity.