Psychological Bulletin
2017, Vol. 143, No. 6, 565–600

0033-2909/17/$12.00 http://dx.doi.org/10.1037/bul0000098

Variables Associated With Achievement in Higher Education: A Systematic Review of Meta-Analyses

Michael Schneider and Franzis Preckel
University of Trier

The last 2 decades witnessed a surge in empirical studies on the variables associated with achievement in higher education. A number of meta-analyses synthesized these findings. In our systematic literature review, we included 38 meta-analyses investigating 105 correlates of achievement, based on 3,330 effect sizes from almost 2 million students. We provide a list of the 105 variables, ordered by the effect size, and summary statistics for central research topics. The results highlight the close relation between social interaction in courses and achievement. Achievement is also strongly associated with the stimulation of meaningful learning by presenting information in a clear way, relating it to the students, and using conceptually demanding learning tasks. Instruction and communication technology has comparably weak effect sizes, which did not increase over time. Strong moderator effects are found for almost all instructional methods, indicating that how a method is implemented in detail strongly affects achievement. Teachers with high-achieving students invest time and effort in designing the microstructure of their courses, establish clear learning goals, and employ feedback practices. This emphasizes the importance of teacher training in higher education. Students with high achievement are characterized by high self-efficacy, high prior achievement and intelligence, conscientiousness, and the goal-directed use of learning strategies. Barring the paucity of controlled experiments and the lack of meta-analyses on recent educational innovations, the variables associated with achievement in higher education are generally well investigated and well understood. By using these findings, teachers, university administrators, and policymakers can increase the effectivity of higher education.

Keywords: academic achievement, meta-analysis, tertiary education, instruction, individual differences

Supplemental materials: http://dx.doi.org/10.1037/bul0000098.supp

Higher education enhances the well-being of individuals and countries. In most industrialized countries across the world, close to 40% of the 25- to 34-year-old citizens have completed tertiary education (Organization for Economic Co-operation and Development [OECD], 2014). Persons with a degree in higher education tend to have better results in adult literacy tests, a lower chance of unemployment, and better health than their peers (Groot & Maassen van den Brink, 2007). At least partly, these are causal effects of education rather than mere correlates. For individual persons, the long-term return on investment (i.e., the benefits minus the costs) for having a tertiary degree instead of just an upper-secondary degree ranges between $110,000 and $175,000 (OECD, 2012). Society as a whole also invests in and profits from higher education. This social return on investment is estimated at $91,000 per student for the OECD countries. Thus, effective higher education brings competence and financial, career, health, and other benefits for individuals and countries.

A key question in the design of effective higher education concerns the sources of students' academic achievement. Which characteristics of students, teachers, and instruction are associated with higher learning outcomes than others? In our study, we use this definition of academic achievement: ". . . performance outcomes that indicate the extent to which a person has accomplished specific goals that were the focus of activities in instructional environments, specifically in school, college, and university [. . .] Among the many criteria that indicate academic achievement, there are very general indicators such as procedural and declarative knowledge acquired in an educational system [and] more curricular-based criteria such as grades or performance on an educational achievement test" (Steinmayr, Meißner, Weidinger, & Wirthwein, 2014). For higher education as well as for education in general, practitioners and researchers in the learning sciences discuss, for example, how strongly achievement is affected by social interaction and student-directed activity versus the mere presentation of content by teachers, by assessment practices, by classroom versus online learning, by prior knowledge and intelligence, and by the students' learning strategies, motivation, personality, and personal background (e.g., Biggs & Tang, 2011; Perry & Smart, 2007; Richardson, Abraham, & Bond, 2012; Schneider & Mustafić, 2015; Schwartz & Gurung, 2012). However, no single empirical study can conclusively evaluate the effectivity of these possible influences on achievement. For example, when an empirical study finds that group work leads to higher learning gains than a lecture, how can researchers know to which extent this conclusion can be generalized beyond the circumstances of that specific study to other content areas, academic disciplines, programs, age groups, institutions, and teachers? Meta-analyses provide a solution to this problem by using math-

This article was published Online First March 23, 2017.

Michael Schneider and Franzis Preckel, Psychology Department, University of Trier.

Correspondence concerning this article should be addressed to Michael Schneider, Psychology Department, University of Trier, Division I, 54286 Trier, Germany. E-mail: m.schneider@uni-trier.de

ematical models to combine the standardized effect sizes obtained in a wide range of empirical studies (Hedges, 1982; Hedges & Vevea, 1998).

Meta-Analyses and Standardized Effect Sizes

By averaging the effect sizes from studies conducted under different circumstances (e.g., in different disciplines or with different teachers), meta-analyses lead to more precise estimates of the true effect sizes, specify the variability and the range of effect sizes across studies, and allow testing for moderator variables, that is, characteristics that explain why the effect sizes are systematically higher in some studies than in others (Rosenthal & DiMatteo, 2001). A meta-analysis usually focuses on a particular instructional method, student characteristic, or teacher characteristic. In the current study, we provide a systematic review of meta-analyses on the variables associated with achievement in higher education. This allows for comparisons of the relative importance of a wide range of variables for explaining academic achievement in higher education, which can inform researchers, teachers, and policymakers.

The results of original empirical studies with different measures can be synthesized in a meta-analysis by considering a standardized index of the effect size instead of the absolute values that were measured. Many studies use Cohen's d as the standardized effect size index, which is usually computed as (mean value of Group A − mean value of Group B)/pooled within-group standard deviation (see Borenstein, 2009; Hattie, 2009, for more detailed explanations). Many other standardized effect size indices, such as the Pearson correlation coefficient r or the proportion of explained variance η^2, are a function of Cohen's d and can easily be converted to this metric (Rosenthal & DiMatteo, 2001).

There are established criteria for how meta-analyses should be conducted and reported (e.g., APA, 2008; Moher, Liberati, Tetzlaff, & Altman, 2009; Shea et al., 2009). The methods and results should be documented in a way that makes it explicit to the reader how the effect sizes were selected, controlled for possible biases, and analyzed. They should document how heterogeneous the obtained findings were and whether part of this variability could be explained by moderator analyses. This information helps researchers evaluate the quality of a meta-analysis and can guide their replication attempts.

Previous Reviews of Meta-Analyses on Achievement

To our best knowledge, no review of meta-analyses on achievement in higher education has been published so far. However, there are at least three such reviews for K–12 school education or for education in general. John Hattie's (2009) book *Visible Learning* reported the most recent and most comprehensive of these reviews. Hattie synthesized the results of about 800 meta-analyses, summing up 52,637 original empirical studies with an estimated total of 236 million participating students. This resulted in a rank-ordered list of 138 variables and the strengths of their associations with achievement (Hattie, 2009, Appendix B).

Hattie noticed that almost all effect sizes were positive, indicating that almost every teaching method led to some learning gains. However, this by no means implies that the choice of the teaching method is irrelevant to achievement and that "anything goes" in teaching. Quite the opposite holds true. The effect sizes varied strongly, indicating that some teaching methods were associated with much greater academic achievement than others. The average effect size for all meta-analyses was close to Cohen's $d = 0.4$. Hattie thus suggested using $d = 0.4$ as the hinge point to differentiate between variables with below or above average effect sizes. The variables whose effect sizes are greater than 0.4 should gain particular attention in the design of learning environments. However, instructional variables with effect sizes smaller than 0.4 can still improve instruction considerably, for example, when their implementation costs little time and money.

Many of the effect sizes that were greater than $d = 0.4$ on Hattie's list were related to what Hattie calls visible teaching and learning, that is, explicit and challenging learning goals, feedback from teachers to learners and vice versa, students actively participating in the learning process, and teachers seeing the learning process through their students' eyes and deliberately trying to improve it (Hattie, 2009, p. 22). Hattie concluded that learning was less effective when teachers and students mindlessly followed routines in the classroom but was more effective when the learning process was made visible, reflected on, and deliberately shaped by teachers and students. These results confirm those of an earlier review of more than 100 meta-analyses (Kulik & Kulik, 1989), which found many moderately strong, positive effect sizes and stressed the importance of clearly defined learning tasks, direct teaching, explicit success criteria, and feedback.

A synthesis of the results of 91 meta-analyses with additional content analyses of qualitative literature reviews and expert ratings concluded that "direct influences like classroom management affect student learning more than indirect influences such as policies" (Wang, Haertel, & Walberg, 1993b, p. 74). The authors used their data sources to create a rank order of six broad groups of variables (Wang, Haertel, & Walberg, 1993a), according to the strength of their association with student achievement. The strongest association was found for (a) student aptitude followed by (b) classroom instruction and climate, (c) home and community educational contexts, (d) curriculum design and delivery, (e) school demographics and organization, and (f) state and district characteristics. As shown by this rank order, proximal variables, which directly capture what is happening in the teaching situation and in the students' minds, are generally more closely associated with achievement than distal variables, which involve the wider context of learning and only indirectly affect what is done and thought in classrooms. This finding supports Hattie's (2009) conclusion that, in addition to student aptitude, what teachers and their students do in learning situations is the most important determinant of achievement.

How Different is Higher Education?

Notwithstanding their impressive merits, previous reviews of the meta-analyses on academic achievement are limited in terms of not reporting separate results for higher education. Some substantial differences between higher education, on one hand, and primary and secondary education, on the other hand, raise the question about the expected degree of similarity among the effect sizes for the three levels. As described in the International Standard Classification of Education (UNESCO, 2012) and in the European Qualifications Framework (European Commission, 2008), primary and secondary education aim at providing students with broad sets of basic knowledge and skills that are important in many areas in later life. In contrast, tertiary education aims at equipping students with advanced knowledge in a specific content domain, along with general competencies in several areas, such as management of complex projects, decision making in unpredictable work contexts, and taking responsibility for the professional development of others. Universities and colleges are more selective than primary, middle, and high schools. University and college students have a longer learning history, more accumulated academic knowledge and skills, and more experience with formal education than school students. Lecture classes in higher educational institutions tend to be larger than school classes. Given these and further differences among primary, secondary, and tertiary education, a review of the meta-analyses on achievement in higher education can complement the existing reviews with their focus on school learning or learning in general and can provide more exact estimates of effect sizes, specifically for colleges and universities.

Debates and Open Questions in Research on Higher Education

A central open question in research on higher education is to what extent instructional methods impact achievement. Previous reviews of meta-analyses found that instructional methods and their implementation in the classroom were among the most important predictors of K–12 student achievement (Hattie, 2009; Kulik & Kulik, 1989). However, students in higher education are a highly select group. To qualify for higher education, a student should have been successful in the K–12 school system. Thus, it is safe to assume that students in higher education have, on average, better intelligence, school achievement, learning strategies, and self-regulation strategies than the overall population (Credé & Kuncel, 2008; Richardson et al., 2012; Sackett, Kuncel, Arneson, Cooper, & Waters, 2009). This mental flexibility might help students in higher education to profit from any instructional method and learning material. Teachers in higher education are usually researchers and experts in the fields of the courses they are teaching. This high degree of domain-specific competence might suffice for them to deliver the courses effectively, irrespective of the teaching methods they use (cf. Benassi & Buskist, 2012; Feldman, 1989).

A second debate in higher education revolves around the role of information and communication technology, particularly about whether online instruction should complement or maybe even replace classroom instruction. Whereas some authors suggested that technology was merely a vehicle that would transport content without changing it (Clark, 1994), others claimed that technology had the potential to completely revolutionize higher education because it would make teaching and learning materials freely available worldwide, could provide quick individualized feedback, and would help people communicate (Barber, Donnelly, & Rizvi, 2013; Pappano, 2012). Still other researchers took a middle ground by asking how the costs of learning technology compare to its benefits and how technology would need to be designed so that it would improve learning (e.g., Kirschner & Paas, 2001; McAndrew & Scanlon, 2013).

A third debate concerns the optimal level of social interaction and student activity in courses (Cherney, 2008). There is a longstanding agreement on the effectivity of the questions asked by a teacher (Redfield & Rousseau, 1981), as well as collaborative learning arrangements (Slavin, 1983). However, it is debated whether lectures and similar teacher-centered forms of instruction, where students spend most of the time listening, are effective and remain a relevant form of instruction (Bligh, 2000). There is also disagreement on the potentials of student projects and similar inquiry-based forms of learning, which require a lot of social interaction and self-directed student activity (Helle, Tynjälä, & Olkinuora, 2006). In these projects, students search and integrate information from different sources, develop and choose from alternative approaches for solving a problem, implement the selected approach, and present the solution, while simultaneously developing and following an adequate time schedule and structuring their group communication. This process provides students with many learning opportunities (Hmelo-Silver, 2004). However, it has been argued that even students with high prior knowledge are usually unable to simultaneously keep track of and perform all of the subtasks involved in project-based learning, so that this is ineffective compared to more teacher-directed instructional methods (Kirschner, Sweller, & Clark, 2006; Mayer, 2004).

Less controversial but nonetheless productive areas of research on achievement in higher education investigate the role of assessment practices (Boud & Falchikov, 2007), student personality (Poropat, 2009), self-regulated learning strategies (Pintrich & Zusho, 2007), multimedia learning (Lamb, 2006), and presentation techniques (Clark, 2008; Levasseur & Sawyer, 2006). These variables are discussed in different literature streams that are rarely brought together. Thus, little is known about how the strengths of their relationships with achievement compare with one another.

The Current Study

In sum, despite the wealth of empirical findings on the variables associated with achievement in higher education, these results have not yet been reviewed and synthesized comprehensively. Against this background, we conducted a systematic literature review of the meta-analyses on the variables associated with achievement in higher education. By compiling the results of meta-analyses, each averaging the effect sizes of the variables across several single studies in a specific subdomain of research, our review gives an overview of a huge range of key findings. However, there is a price to pay for this. Many details of the original empirical studies are lost when they are synthesized in meta-analyses, and not all details of each meta-analysis can be reported in the review. Therefore, it is important to explicitly state the goals and the limitations of our endeavor. We aim at giving researchers and practitioners a broad overview and a general orientation of the variables associated with achievement in higher education. We only included meta-analyses in our review, because they aggregate findings from many empirical studies conducted in different programs, academic disciplines, higher educational institutions, and countries leading to a better estimation precision and generalizability compared with single empirical studies. We go beyond meta-analyses by reporting and comparing the effect sizes from a broad range of literature streams in the learning sciences. Our approach is limited in that we cannot report and discuss detailed information on the measures, samples, or moderating influences separately for all 105 variables included in our review. We refer readers who are interested in these important issues to the meta-analyses listed in Table 1 and to the single studies cited therein. We do not aim to replace these more detailed studies but

Table 1
105 Variables Ordered by the Strength of Their Association With Achievement in Higher Education (Cohen's d)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
1	Instruction	Assessment	Student peer-assessment	Peers grade a student's achievement in addition to the teacher-given grade (high effect size indicates high similarity)	—	3,266	56	1.91	95	yes	n/a	Falchikov and Goldfinch (2000)
2	Student	Motivation	Performance self-efficacy	"Perceptions of academic performance capability" (p. 356)	—	1,348	4	1.81 [1.42, 2.34]	71	no	n/a	Richardson, Abraham, and Bond (2012)[b]
3	Instruction	Stimulating meaningful learning	Teacher's preparation/ organization of the course	e.g., "The instructor was well prepared for each day's lecture. [. . .] The instructor planned the activities of each class period in detail" (p. 633).	—	—	28	1.39	n/a	no	n/a	Feldman (1989)
4	Instruction	Presentation	Teacher's clarity and understandableness	e.g., "The instructor interprets abstract ideas and theories clearly. [. . .] The instructor makes good use of examples and illustrations to get across difficult points" (p. 633).	—	—	32	1.35	n/a	no	n/a	Feldman (1989)
5	Student	Motivation	Grade goal	"Self-assigned minimal goal standards (in this context, GPA)" (p. 357)	—	2,670	13	1.12 [0.93, 1.35]	74	no	n/a	Richardson et al. (2012)[b]
6	Student	Strategies	Frequency of class attendance	Proportion of attended sessions in class	—	21,164	68	0.98	n/a	no	n/a	Credé, Roch, and Kieszczynka (2010)
7	Student	Intelligence and prior achievement	High school GPA	High school grade point average	—	34,724	46	0.90 [0.77, 1.04]	96	no	n/a	Richardson et al. (2012)[b]
8	Instruction	Assessment	Student self-assessment	Students grade their own achievement in addition to teacher-given grades (high effect size indicates high similarity).	—	—	45	0.85	0	no	n/a	Falchikov and Boud (1989)
9	Instruction	Presentation	Teacher's stimulation of interest in the course and its subject matter	e.g., "The instructor puts materials across in an interesting way; the teacher stimulated intellectual curiosity" (p. 633).	—	—	20	0.82	n/a	no	n/a	Feldman (1989)

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
10	Student	Intelligence and prior achievement	Admission test results	Scores on the American College Test (ACT), Scholastic Aptitude Test (SAT), Preliminary SAT (PSAT), and other standardized college admission tests.	—	17,244	17	0.79 [0.64, 0.96]	97	no	n/a	Sackett, Kuncel, Arneson, Cooper, and Waters (2009)
11	Instruction	Social interaction	Teacher's encouragement of questions and discussion	e.g., "Students felt free to ask questions or express opinions; the teacher appeared receptive to new ideas and the viewpoint of others" (p. 634).	—	—	27	0.77	n/a	no	n/a	Feldman (1989)
11	Instruction	Social interaction	Teacher's availability and helpfulness	e.g., "The instructor was willing to help students having difficulty; the teacher was accessible to students outside of class" (p. 635).	—	—	22	0.77	n/a	no	n/a	Feldman (1989)
13	Instruction	Presentation	Teacher's elocutionary skills	e.g., "The teacher speaks distinctly, fluently, and without hesitation; the teacher varied the speech and tone of his or her voice" (p. 633).	—	—	6	0.75	n/a	no	n/a	Feldman (1989)
13	Instruction	Stimulating meaningful learning	Clarity of course objectives and requirements	e.g., "The purpose and policies of the course were made clear to the student; the teacher clearly defined student responsibilities in the course" (pp. 633–634).	—	—	12	0.75	n/a	no	n/a	Feldman (1989)
13	Student	Strategies	Effort regulation	"Persistence and effort when faced with challenging academic situations" (p. 357)	—	8,862	19	0.75 [0.68, 0.82]	23	no	n/a	Richardson et al. (2012)[b]

(*table continues*)

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
16	Instruction	Social interaction	Open-ended questions	Open-ended questions "require pupils to manipulate information to create and support a response"; close-ended questions "call for verbatim recall or recognition of factual information" (p. 237).	Close-ended questions	—	14	0.73	n/a	no	n/a	Redfield and Rousseau (1981)
17	Instruction	Stimulating meaningful learning	Teacher relates content to students	New information is presented in a way that explicitly relates it to the students (self-reference effect).	Teacher presents content without relating it to students	—	60	0.65 [0.58, 0.71]	67	yes	meta	Symons and Johnson (1997)
17	Student	Strategies	Strategic approach to learning	"Task-dependent usage of deep and surface learning strategies combined with a motivation for achievement" (p. 358)	—	2,774	15	0.65 [0.51, 0.81]	70	no	n/a	Richardson et al. (2012)[b]
19	Student	Motivation	Achievement motivation	"One's motivation to achieve success; enjoyment of surmounting obstacles and completing tasks undertaken; the drive to strive for success and excellence" (p. 267)	—	9,330	17	0.64 [0.53, 0.75]	88	no	n/a	Robbins et al. (2004)
20	Instruction	Assessment	Teacher's sensitivity to and concern with class level and progress	e.g., "The teacher was skilled in observing student reactions; the teacher was aware when students failed to keep up in class" (p. 633).	—	—	21	0.63	n/a	no	n/a	Feldman (1989)
21	Student	Motivation	Academic self-efficacy	"General perceptions of academic capability" (p. 356)	—	46,570	67	0.58 [0.46, 0.71]	91	yes	n/a	Richardson et al. (2012)[b]
21	Student	Intelligence and prior achievement	Content of recommendation letter from professor	Letters of recommendation for applications to undergraduate, graduate, and professional schools	—	5,155	6	0.58	n/a	no	n/a	Kuncel, Kochevar, and Ones (2014)

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
23	Instruction	Presentation	Teacher's enthusiasm for subject or teaching	e.g., "The instructor shows interest and enthusiasm in the subject; the teacher communicates a genuine desire to teach students" (p. 633).	—	—	10	0.56	n/a	no	n/a	Feldman (1989)
24	Instruction	Assessment	Quality and fairness of examinations	e.g., "The instructor has definite standards and is impartial in grading; the exams reflect material emphasized in the course" (p. 634).	—	—	25	0.54	n/a	no	n/a	Feldman (1989)
25	Instruction	Assessment	Mastery learning	Students are "required to demonstrate their mastery of each lesson on formal tests before moving on to new material" (p. 265).	Regular instruction	—	86	0.53 [0.45, 0.61]	n/a	yes	meta	Kulik, Kulik, and Bangert-Drowns (1990)
26	Instruction	Stimulating meaningful learning	Intellectual challenge and encouragement of independent thought	e.g., "This course challenged students intellectually; the instructor raised challenging questions and problems" (pp. 634–635).	—	—	7	0.52	n/a	no	n/a	Feldman (1989)
27	Instruction	Technology	Games with virtual reality	"Interactive digital learning environments that imitate a real-life process or situation [with which students can playfully interact and which] provide learners with the opportunities to strategize their moves, test hypotheses, and solve problem" (p. 30).	Traditional instruction or 2D games or no instruction	3,081	13	0.51 [0.25, 0.77]	89	yes	meta	Merchant, Goetz, Cifuentes, Keeney-Kennicutt, and Davis (2014)
27	Instruction	Social interaction	Small-group learning	Small groups of two to 10 students work together toward a common goal.	Individual or whole-class learning	3,472	49	0.51	49	yes	n/a	Springer, Stanne, and Donovan (1999)
29	Instruction	Extracurricular training	Academic skills training	"Interventions, which directly target the skills and knowledge deemed necessary for students to successfully perform in college" (p. 1167)	Various	2,130	17	0.48 [0.31, 0.65]	54	yes	n/a	Robbins, Oh, Le, and Button (2009)

(*table continues*)

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
30	Instruction	Stimulating meaningful learning	Conceptually oriented tasks	Tasks that "elicit students' level of understanding of key science concepts, identify students' misconceptions, [. . . and] engage students with real-world problems in creative ways that reflect a conceptually integrated understanding of the content" (p. 1269)	Other tasks	—	9	0.47	n/a	no	n/a	Ruiz-Primo, Briggs, Iverson, Talbot, and Shepard (2011)
30	Instruction	Assessment	Nature, quality, and frequency of feedback from the teacher to students	e.g., "The teacher gave satisfactory feedback on graded material; criticism of papers was helpful to students" (p. 634).	—	—	20	0.47	n/a	no	n/a	Feldman (1989)
30	Student	Intelligence and prior achievement	Intelligence	Cognitive ability as assessed by standardized tests	—	17,588	26	0.47	n/a	no	n/a	Poropat (2009)
30	Instruction	Social interaction	Teacher's concern and respect for students; friendliness	e.g., "The teacher was friendly toward all students; the teacher took students seriously" (p. 635).	—	—	11	0.47	n/a	no	n/a	Feldman (1989)
30	Student	Personality	Conscientiousness	"Dependability and will to achieve" (p. 322); the tendency to be organized, achievement-focused, disciplined, and industrious	—	32,887	92	0.47	n/a	no	n/a	Poropat (2009)
35	Instruction	Stimulating meaningful learning	Problem-based learning for skill acquisition	Actively solving relatively complex authentic problems in small groups supervised by a teacher or a tutor	Conventional instruction for skill acquisition	—	17	0.46 [0.40, 0.52]	72	yes	n/a	Dochy, Segers, Bossche, and Gijbels (2003)
36	Instruction	Extracurricular trainings	Self-management training programs	"Programs mainly aimed at improving critical skills for effective emotional and self-regulation (e.g., anxiety reduction, desensitization, and stress management/ prevention programs)" (p. 1165)	Various	2,155	28	0.44 [0.22, 0.67]	73	no	n/a	Robbins et al. (2009)

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
37	Instruction	Technology	Simulations with virtual reality	"Simulations are interactive digital learning environments that imitate a real-life process or situation. Simulations allow learners to test their hypotheses of the effects of input variables on the intended outcomes" (p. 30).	Traditional instruction, 2D simulations, or no instruction	2,553	29	0.41 [0.18, 0.64]	85	yes	meta	Merchant et al. (2014)
37	Student	Strategies	Time/study management	"Capacity to self-regulate study time and activities" (p. 357)	—	5,847	7	0.41 [0.25, 0.57]	69	no	n/a	Richardson et al. (2012)[b]
37	Student	Context	Financial support	"Extent to which students are supported financially by an institution" (p. 267)	—	6,849	5	0.41 [0.30, 0.53]	90	no	n/a	Robbins et al. (2004)
37	Student	Strategies	Peer learning	"Tendency to work with other students in order to facilitate one's learning" (p. 357)	—	1,137	4	0.41 [0.03, 0.83]	90	no	n/a	Richardson et al. (2012)[b]
37	Student	Strategies	Learning strategy: organization	"Capacity to select key pieces of information during learning situations" (p. 357)	—	5,410	6	0.41 [0.19, 0.64]	70	no	n/a	Richardson et al. (2012)[b]
42	Instruction	Presentation	Spoken explanation of visualizations (modality effect)	For example, a diagram explained orally as opposed to a diagram with a written explanation	Written explanation of visualizations	—	51	0.38 [0.24, 0.53]	n/a	yes	single	Reinwein (2012)
43	Instruction	Presentation	Animations	Dynamic visualizations, including video-based or computer-based animations and schematic, rather simple, rather realistic, or photo-realistic animations	Static pictures	—	76	0.37 [0.25, 0.49]	82	yes	single	Höffler and Leutner (2007)
43	Student	Strategies	Concentration	"Capacity to remain attentive and task focused during academic tasks" (p. 358)	—	6,798	12	0.37 [0.31, 0.42]	0	no	n/a	Richardson et al. (2012)[b]

(*table continues*)

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
45	Student	Motivation	Academic goals	"One's persistence with and commitment to action, including general and specific goal-directed behavior, in particular, commitment to attaining the college degree; one's appreciation of the value of college education" (p. 267)	—	17,575	34	0.36 [0.31, 0.42]	79	no	n/a	Robbins et al. (2004)
45	Instruction	Stimulating meaningful learning	Concept maps	"A concept map can be regarded as a type of graphic organizer that is distinguished by the use of labeled nodes denoting concepts and links denoting relationships among concepts" (p. 413).	No concept maps	2,496	38	0.36 [0.28, 0.44]	44	yes	meta	Nesbit and Adesope (2006)
47	Instruction	Technology	Intelligent tutoring systems	"Intelligent tutoring systems (ITS) are computer-assisted learning environments. [. . .] They are designed to follow the practices of expert human tutors [and] adjust and respond to learners with tasks or steps suited to the learners' individual [. . .] needs" (p. 331).	Alternative forms of instruction or no instruction	—	37	0.35 [0.24, 0.36]	36	yes	meta	Steenbergen-Hu and Cooper (2014)
47	Student	Strategies	Help seeking	"Tendency to seek help from instructors and friends when experiencing academic difficulties" (p. 357)	—	2,057	8	0.35 [0.21, 0.48]	57	no	n/a	Richardson et al. (2012)[b]
47	Student	Personality	Emotional intelligence	"Capacity to accurately perceive emotion in self and others" (p. 356)	—	5,024	14	0.35 [0.26, 0.43]	33	no	n/a	Richardson et al. (2012)[b]
47	Student	Personality	Need for cognition	"A general tendency to enjoy activities that involve effortful cognition" (p. 356)	—	1,418	5	0.35 [0.05, 0.66]	86	no	n/a	Richardson et al. (2012)[b]

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
51	Instruction	Assessment	Testing aids	"Testing aids' include the use of student-prepared test notes, or crib sheets, as well as the use of textbooks for open-textbook testing conditions" (p. 430).	No testing aids	3,145	35	0.34 [0.15, 0.53]	86	yes	n/a	Larwin, Gorman, and Larwin (2013)
52	Instruction	Technology	Blended learning	"Instructional conditions in which at least 50% of total course time is face-to-face [classroom instruction,] and students working online outside of the classroom spend the remainder of [the] time [. . .] online" (p. 91).	Classroom instruction only	—	117	0.33 [0.26, 0.41]	69	yes	n/a	Bernard, Borokhovski, Schmid, Tamim, and Abrami (2014)
52	Instruction	Extracurricular training programs	Academic motivation training	Training programs on self-motivation strategies that take place outside of the students' regular classes	No academic motivation training	3,720	17	0.33 [0.26, 0.40]	n/a	yes	n/a	Wagner and Szamosközi (2012)
54	Student	Strategies	Critical thinking	"Capacity to critically analyze learning material" (p. 357)	—	3,824	9	0.32 [0.25, 0.40]	0	no	n/a	Richardson et al. (2012)[b]
54	Student	Strategies	Time spent studying	Amount of time spent studying	—	17,242	50	0.32	n/a	no	n/a	Credé and Kuncel (2008)
54	Student	Strategies	Academic-related skills	"Cognitive, behavioral, and affective tools and abilities necessary to successfully complete task, achieve goals, and manage academic demands" (p. 267)	—	16,282	33	0.32 [0.23, 0.42]	76	no	n/a	Robbins et al. (2004)
54	Student	Motivation	Academic intrinsic motivation	"Self-motivation for and enjoyment of academic learning and tasks" (p. 356)	—	7,414	22	0.32 [0.21, 0.44]	83	yes	n/a	Richardson et al. (2012)[b]
58	Student	Personality	Locus of control	"Perceived control over life events and outcomes" (p. 356)	—	2,126	13	0.30 [0.12, 0.49]	78	no	n/a	Richardson et al. (2012)[b]

(*table continues*)

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
59	Student	Context	Social involvement	"The extent that students feel connected to the college environment; the quality of students' relationships with peers, faculty, and others in college; the extent that students are involved in campus activities" (p. 267)	—	15,955	33	0.29 [0.22, 0.35]	85	no	n/a	Robbins et al. (2004)
60	Student	Strategies	Learning strategy: metacognition	"Capacity to self-regulate comprehension of one's own learning" (p. 357)	—	6,205	9	0.28 [0.12, 0.45]	77	no	n/a	Richardson et al. (2012)[b]
60	Instruction	Extracurricular training programs	Training in study skills	"Interventions outside the normal teaching context [. . .] aimed at enhancing motivation, mnemonic skills, self-regulation, study-related skills such as time management, and even general ability itself [. . .]" (pp. 98–99)	No training in study skills	—	103	0.28 [0.23, 0.33]	89	yes	meta	Hattie, Biggs, and Purdie (1996)
60	Student	Motivation	Performance goal orientation	"Achievement striving to demonstrate competence relative to others" (p. 357)	—	18,366	60	0.28 [0.22, 0.35]	73	no	n/a	Richardson et al. (2012)[b]
60	Student	Strategies	Learning strategy: elaboration	"Capacity to synthesize information across multiple sources" (p. 357)	—	8,006	12	0.28 [0.15, 0.42]	84	no	n/a	Richardson et al. (2012)[b]
64	Student	Personality	Optimism	"General beliefs that good things will happen" (p. 356)	—	1,364	6	0.26 [0.13, 0.40]	33	no	n/a	Richardson et al. (2012)[b]
64	Instruction	Presentation	Spoken and written words	e.g., An oral presentation with PowerPoint slides as opposed to an oral presentation without any slideshow.	Spoken words only	1,693	28	0.26 [0.16, 0.36]	n/a	yes	n/a	Adesope and Nesbit (2012)
64	Instruction	Stimulating meaningful learning	Advance organizer	Information (e.g., text or diagrams) presented in the beginning of the learning phase that helps learners to better organize and interpret new content during learning.	No advance organizer	—	40	0.26 [0.08, 0.44]	n/a	no	n/a	Luiten, Ames, and Ackerson (1980)

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
64	Student	Context	Academic integration	"Perceived support from professors" (p. 358).	—	13,755	11	0.26 [0.12, 0.41]	93	no	n/a	Richardson et al. (2012)[b]
68	Student	Context	Socio-economic status	Socioeconomic status (SES) as defined by "father's years of education, mother's years of education, and family income" (p. 4).	—	155,191	41	0.25 [0.21, 0.29]	90	no	n/a	Sackett et al. (2009)
69	Student	Context	Goal commitment	"Commitment to staying at university and obtaining a degree" (p. 358).	—	13,098	10	0.24 [0.09, 0.40]	92	no	n/a	Richardson et al. (2012)[b]
69	Student	Context	Institutional commitment	"Students' confidence of and satisfaction with their institutional choice; the extent that students feel committed to the college they are currently enrolled in; their overall attachment to college" (p. 267).	—	5,775	11	0.24 [0.17, 0.32]	74	no	n/a	Robbins et al. (2004)
69	Student	Personality	Self-esteem	"General perceptions of self-worth" (p. 356).	—	4,795	21	0.24 [0.16, 0.32]	47	yes	n/a	Richardson et al. (2012)[b]
69	Instruction	Assessment	Frequent testing	Students in the frequent testing conditions were tested, on average, 13.6 times per semester (min = 3, max = 50). Students in the infrequent testing conditions were tested, on average, 1.8 times per semester (min = 0, max = 6).	Infrequent testing	—	28	0.24 [0.12, 0.36]	n/a	yes	meta	Bangert-Drowns, Kulik, and Kulik (1991)
69	Student	Motivation	Learning goal orientation	"Learning to develop new knowledge, mastery, and skills" (p. 357).	—	18,315	60	0.24 [0.19, 0.29]	48	yes	n/a	Richardson et al. (2012)[b]
74	Student	Motivation	Control expectation	Personal beliefs about control over life events, as well as generalized and specific control expectancies, assessed as expectancies, locus of control, internal–external attributions, and perceived control	—	2,265	64	0.22	n/a	yes	single	Kalechstein and Nowicki Jr. (1997)

(*table continues*)

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
74	Student	Context	Social support	"Students' perception of the availability of the social networks that support them in college" (p. 267).	—	12,366	33	0.22 [0.17, 0.27]	73	no	n/a	Robbins et al. (2004)
76	Student	Personality	Gender: female	Female vs. male	Gender: male	850,342	131	0.21 [0.17, 0.25]	n/a	yes	n/a	Voyer and Voyer (2014)
77	Instruction	Technology	Group learning with technology	Groups of two to five students were working together during learning. They were either sitting in front of a single computer together or collaborating electronically, each at his or her own computer.	Individual learning with technology	—	178	0.16 [0.12, 0.20]	48	yes	n/a	Lou, Abrami, and d'Apollonia (2001)
78	Student	Personality	Openness	"Imaginativeness, broad-mindedness, and artistic sensibility" (p. 323); the tendency to be open to new feelings, thoughts, and values.	—	28,471	77	0.14	n/a	no	n/a	Poropat (2009)
78	Instruction	Presentation	Students take notes	Note-taking by students during ongoing instruction.	Students take no notes	—	97	0.14 [0.09, 0.20]	69	yes	n/a	Kobayashi (2005)
80	Student	Personality	Agreeableness	"Likability and friendliness" (p. 322); the tendency to be sympathetic, kind, trusting, and co-operative.	—	27,944	75	0.12	n/a	no	n/a	Poropat (2009)
81	Student	Strategies	Learning strategy: rehearsal	"Learning through repetition" (p. 357).	—	3,204	11	0.10 [−0.07, 0.27]	81	no	n/a	Richardson et al. (2012)[b]
81	Student	Personality	General self-concept	"One's general beliefs and perceptions about him/herself that influence his/her actions and environmental responses" (p. 267).	—	9,621	21	0.10	n/a	no	n/a	Robbins et al. (2004)
83	Student	Context	Social integration	"Perceived social integration and ability to relate to other students" (p. 358).	—	19,028	15	0.06 [−0.06, 0.18]	93	no	n/a	Richardson et al. (2012)[b]
83	Student	Context	Institutional integration	"Commitment to the institution" (p. 358).	—	19,773	18	0.06 [−0.10, 0.13]	72	no	n/a	Richardson et al. (2012)[b]
83	Student	Strategies	Deep approach to learning	"Combination of deep information processing and a self (intrinsic) motivation to learn" (p. 358).	—	5,211	23	0.06 [−0.03, 0.15]	60	no	n/a	Richardson et al. (2012)[b]

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
83	Student	Personality	Age	Chronological age	—	42,989	17	0.06 [−0.04, 0.16]	92	no	n/a	Richardson et al. (2012)[b]
83	Student	Personality	Depression	"Low mood, pessimism, and apathy experienced over an extended length of time" (p. 358).	—	6,335	17	0.06 [−0.13, 0.25]	84	no	n/a	Richardson et al. (2012)[b]
88	Instruction	Extracurricular training programs	First-year experience training	"First-year experience can be seen as a type of socialization intervention program. They [generally have a broad] content coverage that aims at improving not only social aspects of the students' life but also their academic skills" (pp. 1165–1166).	Various	2,055	7	0.05 [−0.33, 0.43]	90	no	n/a	Robbins et al. (2009)
88	Instruction	Technology	Online learning	"Online learning [is] learning that takes place entirely or in substantial portion over the Internet" (p. 5).	Face-to-face learning	—	27	0.05	n/a	yes	meta	Means, Toyama, Murphy, and Baki (2013)
90	Student	Motivation	Academic extrinsic motivation	"Learning and involvement in academic tasks for instrumental reasons (e.g., to satisfy others' expectations)" (p. 357).	—	2,339	10	0.00 [−0.14, 0.14]	59	no	n/a	Richardson et al. (2012)[b]
91	Student	Motivation	Pessimistic attributional style	"Perceived control over negative life events and outcomes" (p. 356).	—	1,026	8	−0.02 [−0.27, 0.23]	74	no	n/a	Richardson et al. (2012)[b]
91	Student	Personality	Emotional stability	"Adjustment vs. anxiety" (pp. 322–323); the positive pole of neuroticism, as the tendency to be resilient to negative emotions such as anxiety and depression.	—	28,967	80	−0.02	n/a	no	n/a	Poropat (2009)
91	Student	Personality	Extraversion	"Activity and sociability" (p. 323); the tendency to be friendly, cheerful, sociable, and energetic.	—	28,424	78	−0.02	n/a	no	n/a	Poropat (2009)

(*table continues*)

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
94	Student	Context	Task-related social conflict	"Cognitive-type conflicts [. . .], incompatibilities related to interests or approaches to how work should be done" (p. 117).	No social conflict	3,853	12	−0.13 [−0.16, −0.10]	96	yes	n/a	Poitras (2012)
95	Student	Context	Relationship-related social conflict	"Emotional incompatibilities [. . .] responsible for disputes [. . .] and obstructive or interfering behavior" (p. 117).	No social conflict	3,686	12	−0.21 [−0.24, −0.17]	95	yes	n/a	Poitras (2012)
96	Student	Context	Academic stress	"Overwhelming negative emotionality resulting from academic stressors" (p. 358).	—	941	4	−0.22 [−0.42, −0.03]	48	no	n/a	Richardson et al. (2012)[b]
96	Instruction	Stimulating meaningful learning	Problem-based learning for knowledge acquisition	Actively solving relatively complex authentic problems in small groups supervised by a teacher or tutor.	Conventional instruction for knowledge acquisition	—	18	−0.22 [−0.28, −0.17]	99	yes	n/a	Dochy et al. (2003)
98	Student	Personality	Communication apprehension	"Any feeling of avoidance of interaction with other human beings" (p. 70); synonyms: shyness, speech anxiety, reticence, unwillingness to communicate.	—	8,970	18	−0.23	n/a	yes	n/a	Bourhis and Allen (1992)
99	Student	Motivation	Performance avoidance goal orientation	"Avoidance of learning activities that may lead to demonstration of low ability and achievement" (p. 357).	—	10,713	31	−0.28 [−0.38, −0.19]	79	no	n/a	Richardson et al. (2012)[b]
99	Student	Context	Stress (in general)	"Overwhelming negative emotionality resulting from general life stressors" (p. 358).	—	1,736	8	−0.28 [−0.42, −0.15]	41	no	n/a	Richardson et al. (2012)[b]
101	Instruction	Presentation	Interesting but irrelevant details in a presentation (seductive details effect)	"Seductive details constitute interesting but irrelevant information that are not necessary to achieve the instructional objective" (p. 216).	No interesting but irrelevant details in a presentation	3,535	34	−0.30 [−0.37, −0.22]	84	yes	n/a	Rey (2012)

Table 1 (*continued*)

Rank	Area	Category	Variable	Definition of variable	Compared with	Number of participants	Number of effect sizes	Cohen's *d*, 95% CI in brackets	Heterogeneity I^2	Sign. moderators reported	Support for causality from RCT[a]	Reference
102	Student	Strategies	Academic self-handicapping	"Constructing impediments to performance to protect or enhance one's perceived competence" (p. 744; e.g., effort withdrawal, claiming test anxiety or illness).	—	13,030	25	−0.37 [−0.43, −0.28]	n/a	yes	n/a	Schwinger, Wirthwein, Lemmer, and Steinmayr (2014)
103	Student	Strategies	Surface approach to learning	"Combination of shallow information processing and an extrinsic motivation to learn" (p. 358).	—	4,838	22	−0.39 [−0.55, −0.23]	86	no	n/a	Richardson et al. (2012)[b]
104	Student	Personality	Test anxiety	"Negative emotionality relating to test-taking situations" (p. 357).	—	13,497	29	−0.43 [−0.53, −0.34]	79	no	n/a	Richardson et al. (2012)[b]
105	Student	Strategies	Procrastination	"A general tendency to delay working on tasks and goals" (p. 356).	—	1,866	10	−0.52 [−0.62, −0.42]	5	no	n/a	Richardson et al. (2012)[b]

[a] n/a = Causal relations are not investigated or mentioned. Single = at least one randomized controlled trial (RCT) with evidence of a causal influence on achievement is cited. Meta = Meta-analytical evidence of a causal effect on achievement is presented. [b] The study provided mean effect sizes that were corrected for reliability, but confidence intervals only for noncorrected effect sizes. We estimated the confidence intervals of the corrected effect sizes by centering the noncorrected confidence interval on the corrected effect size.

to complement them. Polanin, Maynard, and Dell (2016) discuss previous reviews of meta-analyses on other topics and give methodological recommendations. We followed these recommendations in our study.

To facilitate the communication and interpretation of the results, we heuristically assigned the 105 variables included in our review to 11 categories. These categories correspond to central areas of educational and psychological research; similar categories were used in previous meta-analyses and reviews of meta-analyses (e.g., Hattie, 2009; Richardson et al., 2012; Wang et al., 1993a). Six of the categories are instruction related, as follows: (a) social interaction, (b) stimulating meaningful learning, (c) assessment, (d) presentation, (e) technology, and (f) extracurricular training. The remaining categories are learner related, as follows: (g) intelligence and prior achievement, (h) strategies, (i) motivation, (j) personality, and (k) context.

Method

Literature Search and Selection Criteria

The details of our search strategy are depicted in Figure 1. In April 2015, we systematically searched the titles, abstracts, and keywords of all articles in the literature database PsycINFO using the search string (*achievement or grades or competence or performance or learning or GPA*) and ("*higher education" or college or university or tertiary*) and limited the results to the meta-analyses published in English in peer-reviewed journals. We explain the reasons for not including grey literature in the next section. In addition to the standardized search, we conducted an exploratory search on GoogleScholar and by scanning the reference lists of relevant reviews, books, book chapters, and articles.

The selection criteria that guided the inclusion of meta-analyses in our systematic review are as follows: (a) The study is a meta-analysis, that is, averaged at least two standardized effect sizes obtained from different samples. (b) The meta-analysis included a measure of achievement as defined in our introduction section. (c) The meta-analysis reported a separate effect size for samples in higher education, or more than 50% of the studies included in the meta-analysis had been conducted with samples in higher education, or the meta-analysis explicitly showed that the effect sizes do not differ between higher education and K–12 school education. (d) Of the found meta-analyses, we only included the largest meta-analysis on each topic, which was usually also the most recent one. (e) The meta-analysis was not explicitly limited to a single subject (e.g., medical education), to a specific subgroup of students (e.g., Latino students), to a single country, or to a single

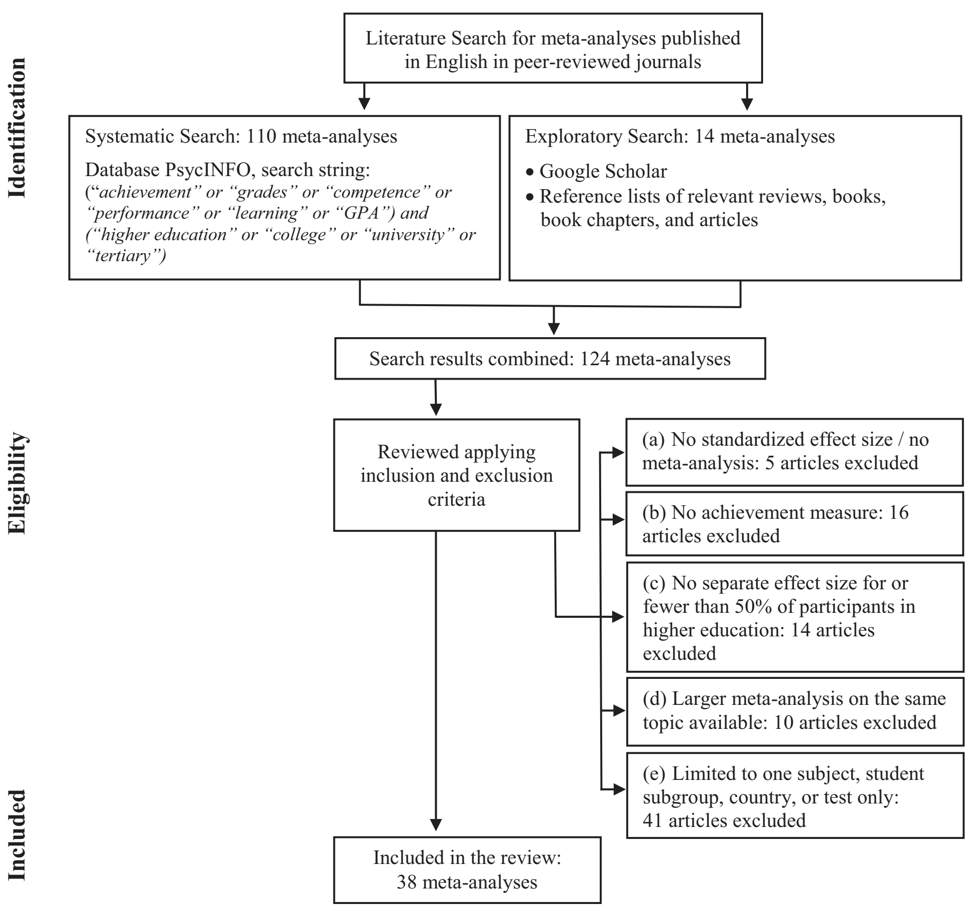

Figure 1. Search strategy and reasons for exclusion.

test (usually with the intention of evaluating its quality and sometimes leading to negative results).

Overall, we found 124 articles. Of these, 86 were discarded based on the selection criteria (see Figure 1). A list with all 124 articles and the 86 reasons for exclusions can be found in the online supplemental material. The results described in this article are based on the remaining 38 meta-analyses, which reported associations between achievement in higher education and 105 other variables.

Extraction of Effect Sizes and Coding

When several meta-analyses on the same topic had been published, Hattie (2009) included all of them and statistically combined their effect sizes. However, this approach makes it difficult to avoid the problem that these meta-analyses might partly include the same studies (Cooper & Koenka, 2012). For example, an older meta-analysis on collaborative learning will include only the old studies on this topic, whereas a newer meta-analysis will cover both the old and the new studies on this subject. When combining the overall results of the two meta-analyses, it is extremely hard to avoid the problem that each older study is included twice, and each newer study is included only once. We thus decided to include only one meta-analysis for each topic, that is, for each variable correlated with achievement. Whenever several meta-analyses on the same topic were available, we selected only the one with the highest number of included empirical studies, which usually was also the most recent meta-analysis (selection criterion d). Thus, in contrast to Hattie's approach, we did not compute any new, combined effect sizes. Instead, we copied the effect sizes from the included meta-analyses.

This also explains why we did not include grey literature in our review. It is sometimes recommended to include grey literature in meta-analyses to reduce publication bias in the effect sizes. However, we conducted a review (of meta-analyses) and not a meta-analysis. As we did not average over effect sizes, including additional results from the grey literature would not have changed the effect sizes we report. It would only have increased the number of variables (i.e., the number of rows) in our main table of results (Table 1). The quality of these additional results would be unknown, because the respective meta-analyses did not pass any peer review. Therefore, we included only meta-analyses published in peer-reviewed international journals in our review. We report in Table 2 which of these meta-analyses checked the results for publication bias.

Each coauthor of the present article coded half of the meta-analyses. Whenever possible, we coded the results of random-effect models, which allow for heterogeneity of the true effect sizes across studies (Hedges & Vevea, 1998). To present all effect sizes in a common metric, we converted Pearson correlations to Cohen's d's by using the formula $d = 2r/\sqrt{1-r^2}$ (Rosenthal & DiMatteo, 2001). Where necessary, we changed the signs so that positive effect sizes always indicated that an instructional method was more effective than standard instruction or that higher values of a continuous variable went along with higher achievement.

Standard errors, 99% confidence intervals, and 90% confidence intervals were transformed into 95% confidence intervals, which we report in brackets. As explained in the Cochrane Handbook of Systematic Reviews of Interventions "The confidence interval describes the uncertainty inherent in [the effect size] estimate, and describes a range of values within which we can be reasonably sure that the true effect actually lies. If the confidence interval is relatively narrow (e.g., 0.70 to 0.80), the effect size is known precisely. If the interval is wider (e.g., 0.60 to 0.93) the uncertainty is greater, although there may still be enough precision to make decisions about the utility of the intervention. Intervals that are very wide (e.g., 0.50 to 1.10) indicate that we have little knowledge about the effect, and that further information is needed" (Schünemann et al., 2011, section 12.4.1). The upper and lower bounds of a confidence interval can be computed as $m \pm z_{1-c/2} * SE_m$, where m is the mean, SE_m is the standard error of the mean, and z is the area under the standard normal distribution for a given level of confidence c (e.g., for 95% confidence intervals, $c = 5\%$ and $z = 1.96$; Belia, Fidler, Williams, & Cumming, 2005). The inverse of this formula allowed us to infer the values of m and SE from 99% and 90% confidence intervals. We then computed the bounds of the 95% confidence interval as $m \pm 1.96 * SE_m$. When a meta-analysis reported neither confidence intervals nor standard errors, we coded and reported the effect size without a confidence interval. Throughtout the manuscript, confidence intervals are on the 95% level and are reported in brackets.

We computed the heterogeneity of effect sizes within each meta-analysis as $I^2 = 100\% * (Q - df)/Q$, where Q is the sum of the squared difference between each effect size and the mean effect size. The degrees of freedom df denote the number of effect sizes minus one (Higgins, Thompson, Deeks, & Altman, 2003). The variable I^2 quantifies the proportion of variance of the effect sizes that cannot be attributed to sampling error and thus indicates the influence of moderator variables. As a variance proportion, I^2 does not quantify the absolute amount of variability in the effect sizes (Borenstein, Higgins, Hedges, & Rothstein, 2017; Rücker, Schwarzer, Carpenter, & Schumacher, 2008).

We classified each effect size as either indicating no effect ($|d| < 0.11$) or a small ($0.11 \leq |d| < 0.35$), medium ($0.35 \leq |d| < 0.66$), or large ($|d| \geq 0.66$) effect. Our cutoff values for these categories were based on Cohen's (1992) suggestion that effect sizes around $d = 0.20$ should be interpreted as small, those around $d = 0.50$ as medium, and those around $d = 0.80$ as large. Effect sizes around $d = 0.00$ indicated no effect. We used the arithmetic means of neighboring values as category boundaries.

Twenty variables were randomly chosen from Table 1. These variables, the characteristics of the respective meta-analyses, and the respective effect sizes were independently coded and converted by both coauthors of this article. The two coders had a perfect interrater reliability, that is, they had 100% agreement in all the coded numbers and characteristics.

Results

Characteristics of the Included Meta-Analyses

The 38 included meta-analyses had been published between 1980 and 2014, with 23 meta-analyses published over the last 10 years (2005 to 2014). They investigated the correlations between 105 variables and achievement in higher education, based on a total of 3,330 effect sizes, and involving an estimated total of 1,920,239 participants. For 30 of the 105 variables, the meta-

Table 2
Methodological Characteristics of the Included Meta-Analyses

Reference	No of fulfilled criteria[a]	Standardized search string	Explicit inclusion and exclusion criteria	Grey literature included	Interrater agreement	Corrected for reliabilities	Corrected for range restriction	Outlier/ sensitivity analysis	Random-effects model	Publication bias: Fail-safe *N*	Publication bias: Trim-and-fill	Publication bias: Other	Publication bias: Evidence of bias[b]	Dependent variables: Standardized achievement tests	Dependent variables: Ad hoc constructed tests	Dependent variables: Teacher-given grades	Dependent variables: Other (e.g., degrees)	Variability of effect sizes: Confidence/ credibility intervals or *SE*	Variability of effect sizes: Heterogeneity index	Variability of effect sizes: Moderator analyses
Total used and reported (%)		26	79	84	37	18	0	34	34	18	16	13	39	58	63	87	63	84	68	95
Adesope and Nesbit (2012)	10	yes	yes	yes	98%	no	no	yes	n/a	yes	no	yes	no	no	yes	no	no	yes	yes	yes
Bangert-Drowns, Kulik, and Kulik (1991)	3	no	no	yes	n/a	no	no	no	n/a	no	no	no	n/a	n/a	n/a	n/a	n/a	yes	no	yes
Bernard, Borokhovski, Schmid, Tamim, and Abrami (2014)	11	yes	yes	yes	92% κ = .84	no	no	yes	yes	yes	yes	no	no	yes	yes	yes	yes	yes	yes	yes
Bourhis and Allen (1992)	3	no	yes	yes	n/a	no	no	no	n/a	no	no	no	n/a	yes	yes	yes	yes	no	no	yes
Credé and Kuncel (2008)	6	no	no	yes	n/a	yes	no	no	yes	no	no	no	n/a	yes	no	yes	yes	yes	yes	yes
Credé, Roch, and Kieszczynka (2010)	9	no	yes	yes	95%	yes	no	no	yes	yes	no	no	n/a	no	no	yes	no	yes	yes	yes
Dochy, Segers, Bossche, and Gijbels (2003)	6	yes	yes	yes	n/a	no	no	no	n/a	no	no	no	n/a	yes	yes	yes	no	yes	yes	yes
Falchikov and Boud (1989)	2	no	no	yes	n/a	no	no	no	n/a	no	no	no	n/a	no	no	yes	no	no	no	yes
Falchikov and Goldfinch (2000)	6	yes	yes	yes	n/a	no	no	yes	n/a	no	no	yes	n/a	no	no	yes	no	no	no	yes
Feldman (1989)	0	no	no	n/a	n/a	no	no	no	n/a	no	no	no	n/a	yes	yes	yes	yes	no	no	no
Hattie, Biggs, and Purdie (1996, pp. 98–99)	6	yes	yes	yes	n/a	no	no	no	n/a	no	no	no	n/a	yes	yes	yes	yes	yes	yes	yes
Höffler and Leutner (2007)	8	no	yes	yes	n/a	no	no	yes	yes	yes	no	no	no	no	yes	no	yes	yes	yes	yes
Kalechstein and Nowicki Jr. (1997)	5	no	yes	yes	n/a	no	no	no	n/a	yes	no	no	no	yes	no	yes	yes	yes	no	yes
Kobayashi (2005)	7	no	yes	yes	84%	no	no	no	n/a	yes	no	no	no	yes	yes	no	yes	yes	yes	yes
Kulik, Kulik, and Bangert-Drowns (1990)	4	no	yes	yes	n/a	no	no	no	n/a	no	no	no	n/a	no	no	yes	no	yes	no	yes
Kuncel, Kochevar, and Ones (2014)	5	no	no	yes	100%	no	no	no	n/a	no	no	no	n/a	no	yes	yes	yes	yes	yes	yes
Larwin, Gorman, and Larwin (2013)	6	no	yes	yes	99%	no	no	no	n/a	no	no	no	n/a	no	no	yes	no	yes	yes	yes
Lou, Abrami, and d'Apollonia (2001)	7	no	yes	yes	93%	no	no	yes	n/a	no	no	no	n/a	yes	yes	yes	yes	yes	yes	yes
Luiten, Ames, and Ackerson (1980)	3	no	no	yes	n/a	no	no	no	n/a	no	no	no	n/a	yes	yes	yes	yes	yes	no	yes
Means, Toyama, Murphy, and Baki (2013)	5	no	yes	yes	n/a	no	no	no	n/a	no	no	no	n/a	yes	yes	yes	yes	yes	yes	yes

Table 2 (*continued*)

Reference	No of fulfilled criteria[a]	Standardized search string	Explicit inclusion and exclusion criteria	Grey literature included	Interrater agreement	Corrected for reliabilities	Corrected for range restriction	Outlier/ sensitivity analysis	Random-effects model	Publication bias: Fail-safe *N*	Publication bias: Trim-and-fill	Publication bias: Other	Publication bias: Evidence of bias[b]	Dependent variables: Standardized achievement tests	Dependent variables: Ad hoc constructed tests	Dependent variables: Teacher-given grades	Dependent variables: Other (e.g., degrees)	Variability of effect sizes: Confidence/ credibility intervals or *SE*	Variability of effect sizes: Heterogeneity index	Variability of effect sizes: Moderator analyses
Merchant, Goetz, Cifuentes, Keeney-Kennicutt, and Davis (2014)	8	no	yes	yes	80–100%	no	no	yes	yes	no	no	no	n/a	yes	yes	yes	yes	yes	yes	yes
Nesbit and Adesope (2006)	8	yes	yes	yes	96%	no	no	yes	n/a	no	no	no	n/a	yes	yes	yes	yes	yes	yes	yes
Poitras (2012)	7	no	yes	yes	n/a	yes	no	yes	n/a	no	no	no	n/a	yes	yes	yes	yes	yes	yes	yes
Poropat (2009)	7	yes	no	yes	n/a	yes	no	no	yes	no	no	no	n/a	no	no	yes	no	yes	yes	yes
Redfield and Rousseau (1981)	2	no	yes	n/a	n/a	no	no	no	n/a	no	no	no	n/a	yes	yes	yes	yes	no	no	yes
Reinwein (2012)	5	no	no	n/a	n/a	no	no	no	yes	no	yes	no	yes	yes	yes	yes	yes	yes	yes	yes
Rey (2012)	6	yes	yes	no	n/a	no	no	no	n/a	yes	no	no	yes	yes	yes	yes	yes	yes	yes	yes
Richardson, Abraham, and Bond (2012)	10	yes	yes	no	$.62 \leq \kappa s \leq 1.00$	yes	no	yes	yes	no	yes	no	in 2 of 50 constructs	no	no	yes	no	yes	yes	yes
Robbins et al. (2004)	8	yes	yes	yes	n/a	yes	no	no	n/a	no	yes	no	no	no	no	yes	no	yes	yes	yes
Robbins, Oh, Le, and Button (2009)	9	no	yes	yes	85%	yes	no	no	yes	no	yes	no	no	no	no	yes	no	yes	yes	yes
Ruiz-Primo, Briggs, Iverson, Talbot, and Shepard (2011)	3	no	yes	no	89%	no	no	no	n/a	no	no	no	n/a	yes	yes	yes	yes	yes	no	no
Sackett, Kuncel, Arneson, Cooper, and Waters (2009, Investigation 2)	6	no	yes	yes	n/a	no	no	no	yes	no	no	no	n/a	no	no	yes	no	yes	yes	yes
Schwinger, Wirthwein, Lemmer, and Steinmayr (2014)	9	no	yes	yes	$.68 \leq \kappa s \leq 1.00$	no	no	yes	yes	no	no	yes	no	yes	yes	yes	yes	yes	yes	yes
Springer, Stanne, and Donovan (1999)	4	no	yes	yes	n/a	no	no	no	n/a	no	no	no	n/a	yes	yes	yes	yes	no	yes	yes
Steenbergen-Hu and Cooper (2014)	9	no	yes	yes	n/a	no	no	yes	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Symons and Johnson (1997)	6	no	yes	yes	n/a	no	no	yes	n/a	no	no	no	n/a	no	yes	no	yes	yes	yes	yes
Voyer and Voyer (2014)	8	no	yes	yes	99%	no	no	yes	yes	no	no	yes	yes	no	no	yes	no	yes	no	yes
Wagner and Szamosközi (2012)	4	no	yes	yes	n/a	no	no	no	n/a	no	no	no	n/a	yes	yes	yes	yes	yes	no	yes

[a] Based on all columns except evidence of publication bias and the four columns for dependent variables. [b] Unlike in all other columns in the table, here we used the 13 meta-analyses that checked for publication bias as 100%, not all 38 meta-analyses.

analyses did not report the total number of included participants (see Table 1). Therefore, we estimated the number of participants by replacing missing values with the median sample size of all included meta-analyses, which was 5,847 (*min* = 941; *max* = 850,342). The median number of the effect sizes included in each meta-analysis was 21 (*min* = 4; *max* = 178). Each meta-analysis reported effect sizes for one to three variables, except for the analyses of Poropat (2009, six variables), Robbins et al. (2004, eight variables), Feldman (1989, 13 variables), and Richardson et al. (2012, 38 variables).

Table 2 presents the methods used and reported in the 38 meta-analyses. The first row of the table summarizes the percentage of the 38 meta-analyses that used and reported a certain method. Whereas most meta-analyses gave detailed descriptions of how they conducted the literature search, only 26% reported the exact search string they used in each database. The majority of the meta-analyses (79%) reported explicit inclusion and exclusion criteria. Most of them (84%) included gray literature, which could help reduce publication bias. Interrater agreement was reported in 37% of the studies. Only 18% of the meta-analyses corrected the effect sizes for the nonperfect reliabilities of the measures (cf. Muchinsky, 1996). No meta-analysis corrected for the range restriction. This restriction of the variance occurs in research on higher education because college and university students are more homogeneous with respect to their achievements, intelligence, socioeconomic status, and so on, compared with the overall population. However, it can be argued that a correction for range restriction is not needed because the authors generalized their results only to university and college students, not to persons outside higher education. Only 34% of the meta-analyses reported that they scanned the results for outliers, which could have biased the results. One third of the meta-analyses (34%) explicitly reported the use of a random-effects model, which allowed for heterogeneity of the effect sizes, for example, due to the influence of moderator variables. Of the 38 meta-analyses, 34% checked for publication bias in some form (Duval & Tweedie, 2000). Of these 13 studies, five (i.e., 39%) found evidence of publication bias and eight found evidence of the lack of this bias. The meta-analyses included in our review were relatively homogeneous in terms of the dependent variables used. Standardized achievement tests were used in 58% of the studies, ad hoc-constructed tests in 63%, teacher-given grades in 87%, and other measures in 63% of the studies. Confidence intervals or standard errors of the average effect sizes were reported in 84%, a heterogeneity index (e.g., I^2 or Q) was given in 68%, and moderator analyses were described in 95% of the meta-analyses.

In Table 2, we list frequently used criteria for evaluating the quality of meta-analyses (cf. Schmidt & Hunter, 2015) and to what extent the included meta-analyses fulfill these criteria. The overall quality of a meta-analysis depends on many details and cannot be reduced to the listed criteria. Therefore, we used the number of criteria fulfilled by each meta-analysis only to heuristically distinguish three groups of studies: meta-analyses fulfilling a low number (i.e., three or less), a medium number, and a high number (i.e., nine or more) of quality criteria. Seven meta-analyses fulfilled nine or more criteria. They investigated verbal redundancy in multimedia learning (Adesope & Nesbit, 2012), blended learning (Bernard, Borokhovski, Schmid, Tamim, & Abrami, 2014), class attendance (Credé et al., 2010), students' psychological correlates of achievement (Richardson et al., 2012), college interventions (Robbins, Oh, Le, & Button, 2009), academic self-handicapping (Schwinger, Wirthwein, Lemmer, & Steinmayr, 2014), and intelligent tutoring systems (Steenbergen-Hu & Cooper, 2014). Seven other meta-analyses only fulfilled three or fewer quality criteria. They analyzed frequent classroom testing (Bangert-Drowns, Kulik, & Kulik, 1991), communication apprehension (Bourhis & Allen, 1992), student self-assessment (Falchikov & Boud, 1989), associations between student ratings of instruction and achievement (Feldman, 1989), advance organizers (Luiten, Ames, & Ackerson, 1980), teacher questions (Redfield & Rousseau, 1981), and undergraduate science course innovations (Ruiz-Primo, Briggs, Iverson, Talbot, & Shepard, 2011). The meta-analysis contributing the highest number of variables to our main results (Richardson et al., 2012, 38 variables) followed a high number of recommendations and was of an excellent methodological quality. The meta-analysis contributing the second-highest number of variables to our main results (Feldman, 1989, 13 variables) did not follow a single of the recommendations in Table 2. It is thus unclear to what extent the findings reported by Feldman are comprehensive and representative for the literature, are distorted by outliers, suffer from publication bias, vary between studies, and are moderated by third variables.

Characteristics of the Included Effect Sizes

Table 1 lists all 105 variables, ordered according to the strength of their association with achievement. A high Cohen's *d* indicates that high values of the respective variables tend to be linked with high student achievement. Positive values indicate a positive association (e.g., more open-ended questions go along with higher student achievement), negative values indicate an inverted association (e.g., lower test anxiety goes along with higher achievement). The effect sizes range from −0.52 to 1.91, with a mean of 0.35, a median of 0.33, and a standard deviation of 0.41. Half of the effect sizes lie between 0.13 and 0.51. The frequency distribution of the 105 effect sizes is displayed in Figure 2. When an effect size was derived by comparing an intervention group with a control group, the control group is described in the sixth column of Table 1. For example, students of teachers mainly using open-ended questions outperformed students of teachers mainly using close-ended questions, by $d = 0.73$ on average. When an effect size was computed by correlating the values of two continuous variables, the corresponding row under the sixth column is empty.

A decrease of a variable with a negative effect size can increase achievement as well as the increase of a variable with a positive effect size can. For example, teachers can increase achievement by decreasing the number of seductive details in their presentations. Seen this way, all variables in Table 1 can be used to make instruction more effective irrespective of the signs of their effect size values. Thus, we also computed the descriptive characteristics of the absolute effect sizes (i.e., without their signs). They range from 0.00 to 1.91, have a mean of 0.42, a median of 0.35, and a standard deviation of 0.33.

Confidence intervals were available for 74 of the 105 effect sizes, as shown in Table 1. Confidence intervals were unavailable for some of the older meta-analyses, as well as for most of the newer meta-analyses that reported credibility intervals instead of confidence intervals. Some meta-analyses reported confidence

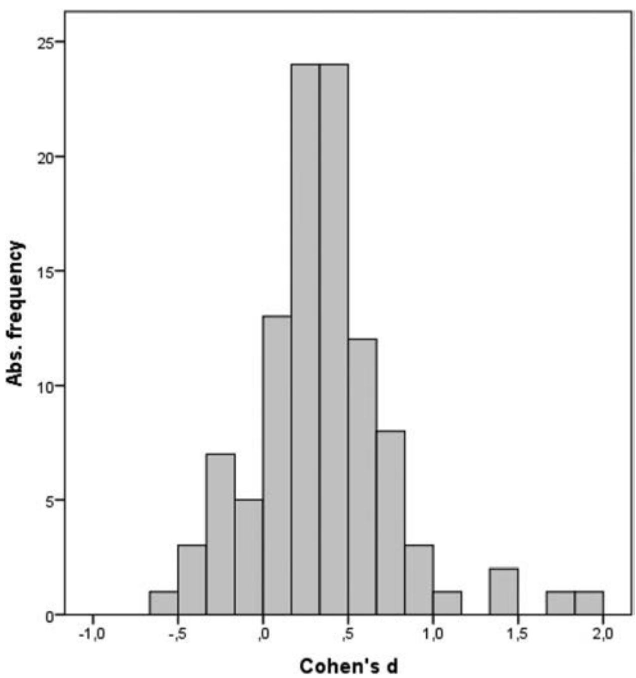

Figure 2. Distribution of the 105 effect sizes for the associations with achievement.

intervals only for the overall average effect sizes but not for moderator analyses. Only five of the 105 confidence intervals were wider than $\Delta d = 0.50$ (for performance self-efficacy, virtual-reality games, peer learning, need for cognition, and first-year experience trainings). The heterogeneity index I^2 was reported for 70 of the 105 effect sizes. The I^2 values ranged from 0% to 99%, with a median of 75%. The effect size d was uncorrelated with the number of effect sizes k, $r = -.002$, $p = .981$, the number of participants N, $r = -.021$, $p = .859$, and the amount of heterogeneity I^2, $r = .021$, $p = .862$, as listed in Table 1.

A few of the meta-analyses explicitly analyzed or discussed whether the effect sizes indicated a causal effect on achievement (see column 12 in Table 1). Most meta-analyses did not provide separate average effect sizes for randomized controlled trials (RCTs), which could yield conclusive evidence of causal relations. Meta-analytical evidence in favor of a causal effect on achievement, that is, a separate mean effect size for RCTs, was presented only for nine of the 105 variables (the teacher relating the content to students, mastery learning, games with virtual reality, simulations with virtual reality, concept maps, intelligent tutoring systems, training in study skills, frequent testing, and online learning). Single RCTs with evidence of a causal effect on achievement were cited for three further variables (spoken explanation of visualizations, animations, and control expectations). No meta-analysis reported empirical evidence against the assumption of a causal effect of the investigated variable on achievement.

Table 3 lists the 11 categories that we used to group the 105 variables. For each category, it summarizes the total numbers of included participants, effect sizes from empirical studies, and investigated variables, along with the percentages of the variables with small, medium–strong, or strong effects. In Table 3, we ordered the variables under the instruction-related categories and the learner-related categories by the combined proportion of medium–strong and strong effects. In the following 11 subsections, we briefly describe the findings for each of the six instruction-related and five learner-related categories.

Instruction Variables

Social interaction. Five variables related to social interaction have been investigated in the meta-analyses (see Table 3). Two of them have medium–strong associations with achievement, and the other three have strong associations. Thus, the social interaction category has a higher proportion of variables with medium–large

Table 3
Absolute Frequencies of Data Points and Percentage of Variables by Effect Size (Ordered by the Combined Frequency of Medium and Large Effects)

	Absolute frequency of data points			% of variables			
	Students[a]	Effect sizes	Variables	No effect	Small effect	Medium effect	Large effect
Overall	1,920,239	3,330	105	12	36	36	15
Instruction variables	208,711	1,595	42	5	26	45	24
Social interaction	26,860	123	5	0	0	40	60
Stimulating meaningful learning	49,272	229	9	0	22	56	22
Assessment	41,493	316	8	0	25	50	25
Presentation	46,157	354	9	0	33	33	33
Technology	29,022	401	6	17	33	50	0
Extracurricular training programs	15,907	172	5	20	40	40	0
Student variables	1,711,528	1,735	63	18	43	30	10
Intelligence and prior achievement	74,711	95	4	0	0	50	50
Strategies	133,757	343	18	11	28	50	11
Motivation	137,880	390	12	17	42	25	17
Personality	1,093,174	694	16	31	44	25	0
Context	272,006	213	13	15	77	8	0

Note. No effect = $|d| < .11$; small effect = $.11 \leq |d| < .35$; medium effect = $.35 \leq |d| < .66$; large effect = $|d| \geq .66$.
[a] Estimated by replacing missing values of a meta-analysis by the median value of all meta-analyses.

and large effect sizes than any other instruction-related category. The five variables in this category are listed in Table 1. The variable with the strongest relation to achievement is teachers' encouragement of questions and discussion (d = 0.77, rank 11 in Table 1). Open-ended questions (d = 0.73, rank 16), such as "Why . . .?," "What is your experience with . . .?," and "What are the disadvantages of . . .?" are more effective than close-ended questions, such as "In which year . . .?" or "Who did . . .?" because open-ended questions require students to explain, elaborate, or evaluate, whereas close-ended questions only require recalling memorized facts. An experiment (Kang, McDermott, & Roediger, 2007) replicated the older, correlational findings of the meta-analysis and yielded evidence for a causal effect of open-ended questions on learning outcomes.

Classes with high-achieving students complement these forms of teacher–student interactions with student–student interactions. On average, small-group learning goes along with higher achievement than individual learning or whole-group learning (d = 0.51, rank 27; Springer, Stanne, & Donovan, 1999). The investigated groups comprised two to 10 persons who solved a task together during a course. Small-group learning is more effective when each learner has individual responsibilities within his or her group, and when the learners can only solve the task through cooperation (Slavin, 1983). Thus, the success of small-group learning critically depends on choosing appropriate tasks. King (1990, 1992) reported examples of effective small-group learning activities in large lecture classes.

The remaining two variables in the social interaction category are teachers' availability and helpfulness (d = 0.77, rank 11), as well as friendliness, concern, and respect for students (d = 0.47, rank 30). These variables indicate the importance of creating an atmosphere where students are comfortable with answering questions, sharing their views, and engaging in social interactions with their teacher and with one another.

Overall, the finding that social interaction is strongly associated with achievement in higher education is consistent with the results of studies on school learning and other formal and informal learning settings (Hattie, 2009; Ruiz-Primo et al., 2011). Part of what makes social interaction so effective is that it requires active engagement, explicit verbalization of a person's own knowledge, perspective taking, and the comparison of arguments and counterarguments (Chi, 2009).

Stimulating meaningful learning. Stimulating meaningful learning is the category with the second highest proportion of medium–high and high effect sizes relating to instruction in Table 3. Seven of the nine variables in this category show medium-large or large effect sizes. As indicated by the nine variables (see Table 1), meaningful learning requires thoughtful preparation and organization of the course (d = 1.39, rank 3), as well as clear course objectives and requirements (d = 0.75, rank 13), so that teachers and students do not mindlessly follow routines but intentionally engage in educational practices to attain their goals. Teachers can also make new content more meaningful by explicitly pointing out how it relates to the students' lives, experiences, and aims (d = 0.65, rank 17). This helps students in perceiving the relevance of the new content, integrating it with prior knowledge in their long-term memory, and memorizing it (Schneider, 2012). Providing intellectual challenges and encouraging independent thought (d = 0.52, rank 26) helps students think through new learning content, elaborate on it, consider its relation to their prior knowledge, and identify its theoretical and practical implications.

Teachers can foster meaningful learning by beginning each lesson with a short description of its contents (advance organizer, d = 0.26, rank 64) and by visualizing abstract relations between constructs in concept maps (d = 0.36, rank 45). As moderator analyses show, concept maps are more effective when they depict central ideas only (d = 0.60, CI [0.40, 0.79]) than when they also show details (d = 0.20, CI [0.02, 0.39]; Nesbit & Adesope, 2006). The so-called conceptually oriented tasks (d = 0.47, rank 30) are useful, because they "elicit students' level of understanding of key science concepts, identify students' misconceptions, [. . . and] engage students with real-world problems in creative ways that reflect a conceptually integrated understanding of the content" (Ruiz-Primo et al., 2011, p. 1269).

Classes can be made more meaningful by using project-based learning arrangements, where groups of students work on complex authentic tasks over extended periods of time and have to structure their own problem-solving process under the supervision of a teacher or a tutor. The students' projects are often similar to scientific research or workplace projects, helping students to see where and how the contents their course can be useful. Project-based learning is more effective than regular lectures and seminars for acquiring practical skills (d = 0.46, rank 35) but less effective for acquiring fact knowledge (d = −0.22, rank 96). Thus, it can be productive to complement project-based learning with lectures, which improve knowledge more strongly than practical skills (Bligh, 2000, p. 5). According to moderator analyses, an entire project-based curriculum (d = 0.31, CI [0.23, 0.40]) has stronger positive effects on skills than a single project-based course (d = 0.19, CI [0.10, 0.27]; Dochy, Segers, Bossche, & Gijbels, 2003). Similar to all forms of small-group learning, project-based learning requires careful preparation and close supervision by teachers. Entirely student-directed project-based learning, the so-called pure discovery learning, is far less effective than standard lectures and seminars (d = −0.38 in a meta-analysis by Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; cf. Mayer, 2004).

Assessment. As shown in Table 3, six of the eight variables in the assessment category have medium–large or large effect sizes. From an educational perspective, the most important function of assessment is to provide learners and teachers with feedback about past progress and their needs for future developments (Hattie & Timperley, 2007). Thus, it is not only summative assessment at the end of an instructional unit that can be effective but also formative assessment at the beginning of or during an ongoing unit (Angelo & Cross, 1993; Nicol & Macfarlane-Dick, 2006).

As shown in Table 1, the highest effect sizes in this category are found for student–peer assessment (d = 1.91, rank 1) and student self-assessment (d = 0.85, rank 8). These two meta-analyses (Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000) compared self-given, peer-given, and teacher-given grades. The high effect sizes indicate that students themselves, their peers, and their teachers all tend to assign similar grades when evaluating achievement. This convergence suggests that grades in higher education are on average relatively objective and reliable. The two meta-analyses report mere correlations and do not imply whether self-assessment or peer-assessment causally affects achievement, which remains an open question.

Clear learning goals and success criteria are strongly associated with achievement in higher education (clarity of course objectives and requirements, d = 0.75, rank 13). Teachers need them for semester and lesson planning, evaluating achievement, and providing individual feedback. Students need them for distinguishing between important and unimportant lesson contents, planning and organizing learning and problem solving, and preparing for exams (cf. Seidel, Rimmele, & Prenzel, 2005). Thus, clear course objectives and requirements have dual functions. They guide lesson planning, as well as assessment, and help integrate these two components of instruction. For this reason, we have assigned this variable to the *stimulating meaningful learning* category in Table 1 but also refer to it here under the Assessment section.

Important for achievement are the teacher's sensitivity to and concern with class level and progress (d = 0.63, rank 20), the quality and fairness of examinations (d = 0.54, rank 24), as well as the nature, quality, and frequency of feedback from the teacher to the students (d = 0.47, rank = 30). When all students are required to demonstrate their mastery of the lesson content on a test before the instruction moves on, this raises their achievement by about half a standard deviation (mastery learning, d = 0.53, rank 25). Allowing for open textbooks or student-prepared notes during exams increases the similarity between the exam situation and real-life problem solving but also makes the exam slightly easier compared with traditional exams (testing aids, d = 0.34, rank 51; specifically, d = 0.40 for student notes and d = 0.26 for textbooks; Larwin, Gorman, & Larwin, 2013). Testing frequently, for example, weekly leads to somewhat higher learning gains than testing infrequently, for instance, only once or twice during a semester (d = 0.24, rank 69).

Presentation. The variables in the presentation category refer to how a teacher delivers the course content to students. Six of nine effect sizes in this category are medium–large or large (see Table 3). The associations with student achievement were strongest for clarity and understandability (d = 1.35, rank 4), the teacher's stimulation of the students' interest (d = 0.82, rank 9), elocutionary skills (d = 0.75, rank 13), and the teacher's enthusiasm about the course or its content (d = 0.56, rank 23).

Teachers can improve their presentations by adhering to several design principles. Presentations that combine spoken with written language (e.g., a talk with a slideshow) are more effective (d = 0.26, rank 64) than solely oral presentations, because the written words focus the learner's attention on key points and aid in memorization. As the moderator analyses show, this method works much better with a few written keywords (d = 0.99) than with written half sentences or full sentences (d = 0.21; Adesope & Nesbit, 2012). Sentences or half sentences distract from the spoken words and place an unnecessarily high load on the verbal working memory. Thus, the popular presentation technique of complementing a teacher's talk with a slideshow is effective, particularly when only a few brief keywords (or a picture) are displayed on each slide.

When diagrams or pictures are presented, they should be complemented by spoken rather than written language (d = 0.38, rank 42). For example, a PowerPoint slide with a diagram on it should be explained orally rather than by some text on the slide. Only in the former case can learners focus all of their visual attention on the figure while listening to the explanation (Ginns, 2005). This effect is stronger for dynamic visualizations (d = 0.82, CI [0.62, 1.03]) than for static ones (d = 0.23, CI [0.11, 0.35]; Reinwein, 2012). Animations visualize processes or movements more effectively than static pictures (d = 0.37, rank 43).

All content and design elements that are unnecessary for achieving the predefined learning goals should be omitted from a presentation because they distract from the content to be learned (Mayer & Moreno, 2003). This is even true for unnecessary design elements that are interesting or decorative (seductive details effect, d = −0.30, rank 101). The seductive details effect is weak for learning without time pressure (d = −0.10, CI [−0.20, 0.00]) and stronger for learning under time pressure (d = −0.66, CI [−0.78, −0.54]; Rey, 2012).

Students' note-taking during presentations has a weak effect on achievement (d = 0.14, rank = 78). However, there is an important moderator to consider. Note-taking is helpful when there are no presentation slides (d = 0.43) but is completely ineffective when there are presentation slides (d = −0.02; Kobayashi, 2005). This finding is plausible because listening and simultaneously writing are easier than listening, looking at slides, and simultaneously writing.

Technology. Information and communication technologies in higher education include online lectures and podcasts, massive open online courses (MOOCs), online learning platforms, Internet discussion forums, instructional videos and simulations, serious games, clickers, social media such as wikis, and so on. Three of the six variables in the technology category have medium–large effect sizes, and there are no large effect sizes. Online learning is about as effective as learning in the classroom (with a difference of d = 0.05, rank 88). In contrast, blended learning, that is, a mix of online and classroom learning, is more effective than classroom learning alone (d = 0.33, rank 52). In their meta-analysis, Bernard et al. (2014) characterized blended learning as "the combination of face-to-face and online learning outside of class" but defined no minimum of course time for online activities in blended learning leaving it unclear whether, for example, uploading presentation slides on an Internet server for the students already counted as blended learning in their analysis. Obviously, blended learning can include a very heterogeneous set of instructional approaches what hampers the interpretation of the results. Moderator analyses show that blended learning is more effective when instructional technology is used to support cognition (d = 0.59, CI [0.38, 0.79]), for example, by visualizing abstract concepts, than when it is applied to support communication, for example, group chats (d = 0.31, CI [0.07, 0.55]; Bernard et al., 2014). In line with this, the benefits of small-group learning are higher when the groups interact face-to-face (d = 0.51, rank 27) as to compared with when they work with technology (d = 0.16, rank 77). One possible reason is that online communication limits the range of social interactions compared to face-to-face meetings (Lou, Abrami, & d′Apollonia, 2001).

Two resource-intensive but comparatively effective technological interventions are games with virtual reality (d = 0.51, rank 27) and interactive virtual-reality simulations of real-world processes (d = 0.41, rank 37). However, both effect sizes have confidence intervals with a width of about 0.5 what indicates a low precision of estimation. As moderator analyses show, the simulations are effective after an instruction on the relevant concepts (d = 0.59) but ineffective as a standalone form of instruction (d = 0.09; Merchant, Goetz, Cifuentes, Keeney-Kennicutt, & Davis, 2014). Playing instructional games individually (d = 0.72) is more effective than playing them in groups (d = 0.00), perhaps because of the competition and group dynamic distraction from the content to be learned (Merchant et al., 2014). Intelligent tutoring systems (d = 0.35, rank 47) are computer learning environments in which artificial intelligence is used to analyze a student's learning process and to provide learning tasks and feedback that continuously adapt to the learner's individual progress. The moderator analyses show that intelligent tutoring systems are considerably more effective than learning without technology or not learning at all (d = 0.86) but are less effective than human tutors (d = −0.25; Steenbergen-Hu & Cooper, 2014).

Four of the six meta-analyses in the technology category tested whether the effect sizes increased over time, what could have been interpreted as positive educational effects of technological advances. The three meta-analyses on small-group learning with technology (Lou et al., 2001), online learning (Means, Toyama, Murphy, & Baki, 2013), and blended learning (Bernard et al., 2014) found no change over time, and the meta-analysis on intelligent tutoring systems (Steenbergen, Craje, Nilsen, & Gordon, 2009) even found a decrease.

Extracurricular training. Extracurricular training programs are offered outside of the students' regular classes to improve general academic skills, such as learning strategies, critical thinking, or self-motivation. Only two of the five variables in this category have a medium–strong effect, and none has a strong effect. The effect of the training in study skills (in general) on achievement is d = 0.28 (rank 60). The most effective form of training in study skills is the academic type "which directly target[s] the skills and knowledge deemed necessary for students to successfully perform in college" (d = 0.48, rank 29; Robbins et al., 2009, p. 1167). It is closely followed by self-management training programs (d = 0.44, rank 36), for example, anxiety reduction, desensitization, and stress management or stress prevention. Training programs in academic motivation are slightly less effective (d = 0.33, rank 52). Finally, seven so-called first-year experience training programs that had been evaluated with 2,055 students had virtually no average effect on achievement (d = 0.05, rank 88). These programs combine general information and orientation with student socialization components and training in academic skills. The confidence interval for this effect size from −0.33 to 0.43 is so wide that no definite conclusions about the effectivity of first-year experience programs can be drawn.

A general limitation of extracurricular training is that its far transfer to tasks that are dissimilar to the training problems is much weaker (d = 0.33) than its near transfer to very similar tasks (d = 0.57, Hattie, Biggs, & Purdie, 1996). Hattie and colleagues concluded "a very strong implication of this is that study skills training ought to take place in the teaching of content rather than in a counseling or remedial center as a general or all-purpose package of portable skills" (Hattie et al., 1996, p. 130). Other literature reviews with different foci reached the same conclusion. Training general academic skills in the context of concrete classes with clear program-related learning goals is usually more effective than extracurricular training with artificial problems (Paris & Paris, 2001; Pressley, Harris, & Marks, 1992; Tricot & Sweller, 2014).

Student Variables

Intelligence and prior achievement. The four variables in the intelligence and prior achievement category all have medium–large or large effect sizes, highlighting the importance of prior achievement and intelligence for achievement in higher education. Of all the person-related categories (see Table 3), the prior achievement and intelligence category shows the strongest relation with achievement.

There are multiple reasons behind the large effect sizes for the grade point average (GPA) in high school (d = 0.90, rank 7) and the admission test results (d = 0.79, rank 10). Prior knowledge supports the acquisition of new knowledge; it aids in future learning (Hambrick, 2003). Prior achievement and knowledge result from prior effort and investment, which have some stability over time (Gottfried, Fleming, & Gottfried, 2001). Past academic success is a reward that enforces further engagement in learning processes (positive reciprocal effects; e.g., Marsh & Martin, 2011). Prior achievement and subsequent achievement are also related because they are both affected by intelligence as a relatively stable personal characteristic (Neisser et al., 1996). However, when directly correlated with achievement, intelligence only shows a medium–large effect (d = 0.47, rank 30), possibly because intelligence tasks tend to be more abstract and generic than admission test tasks.

Finally, when the contents of the professors' recommendation letters for students are rated quantitatively, the results have a medium–strong correlation with achievement (d = 0.58, rank 21). In their meta-analysis, Kuncel, Kochevar, and Ones (2014) found that these letters had incremental validity over traditional predictors for explaining the degree attainment in graduate school. The authors concluded "that [these] letters [might] be able to tap into the much needed motivation and persistence that are difficult to obtain through most admission tools" (Kuncel, Kochevar, & Ones, 2014, p. 105), and they gave recommendations for further improvement of these letters (e.g., focus on noncognitive constructs).

Strategies. The strategies category merges 18 variables related to students' self-regulated learning strategies and approaches to learning. Self-regulated learning strategies are used by learners to systematically and actively attain their personal goals and involve the regulation of cognition, behavior, and affect (Zimmerman & Schunk, 2011). Approaches to learning refer to different strategies students can adopt to learn or to complete a task (e.g., surface vs. deep approach; Marton & Säljö, 1976). Compared with self-regulatory strategies, they describe broader characterizations of learning tendencies (Pintrich, 2004). Eleven of the 18 variables in this category show medium–strong or strong effects.

The frequency of class attendance has the largest effect size in this category (d = 0.98, rank 6). Students who attend more class sessions show significantly better achievement than students with lower attendance rates. This correlational finding does not allow for a causal interpretation (see Credé et al., 2010, for a detailed discussion). However, the empirical results indicate that the frequency of class attendance makes unique contributions to academic achievement beyond prior achievement and personality traits such as conscientiousness. Furthermore, the effect of class attendance has remained constant over the past years. Thus, the increasing frequency of online classes and blended learning (i.e., presentation slides for download) does not diminish the importance of class attendance for achievement. So far, there is hardly any controlled quantitative study that has investigated the effects of mandatory attendance policies, so it is too early to draw conclusions about their usefulness.

The more the students engage in effort regulation (i.e., respond to challenging academic situations with persistence and effort), the higher their achievement is (d = 0.75, rank 13). Furthermore, achievement is higher the more the students employ a strategic approach to learning (d = 0.65, rank 17), that is, use learning strategies in a task-dependent way, combined with their motivation for achievement. To what extent the students use a deep approach to learning, that is, elaborated new content on a deep level, had no systematic relation with achievement (d = 0.06, rank 83). In contrast, a surface approach to learning, that is, the combination of shallow information processing and a focus on external rewards, had a medium–strong negative relation with achievement (d = −0.39, rank 103).

Resource management strategies (i.e., time/study management, peer learning, and help seeking) are positively related to achievement with medium–large effect sizes (d′s between 0.35 and 0.41, see Table 1). Several cognitive and metacognitive strategies (i.e., organization, concentration, critical thinking, metacognition, elaboration, and rehearsal) have small to medium–large effect sizes (with ds between 0.10 and 0.41, see Table 1). In sum, these findings stress the importance of self-regulated learning strategies, particularly the management of effort and time, for students in higher education. This conclusion is also supported by the negative medium-sized effects of maladaptive strategies, indicating a lack of self-regulatory competencies, namely academic self-handicapping (d = −0.37, rank 102) and procrastination (d = −0.52, rank 105).

Motivation. Five of the 12 variables in the motivation category have a medium-large or large effect size. We find a very large effect size for performance self-efficacy (d = 1.81, rank 2). The very wide confidence interval from 1.42 to 2.34 indicates that this is just a rough estimate of the population effect size. Self-efficacy is the belief in one's ability to plan and to execute the skills necessary to produce a certain behavior (Bandura, 1979). Studies conducted outside of higher education show that self-efficacy has a positive causal effect on achievement, which in turn causally affects self-efficacy. Teachers can improve their students' self-efficacy by giving them a sense of achievement in the context of demanding tasks, defining clear learning goals, and setting explicit standards for success (Bandura, 1993). These elements help students recognize their potential for success in class and their means to accomplish it. Part of the correlation between self-efficacy and achievement can also be attributed to the fact that prior achievement and intelligence affect both self-efficacy and achievement. Whereas performance efficacy is always directed at a specific performance goal, academic self-efficacy is the students' more general perception of their academic competence. This variable is less strongly related to achievement (d = 0.58, rank 21), likely because it is easier for students to predict their success in a concrete exam than their general success in academia.

High-achieving students set grade goals for themselves (d = 1.12, rank 5), which are their minimum standards for their target grades. They have a high achievement motivation (d = 0.64, rank 19; Robbins et al., 2004, p. 267). Related to this, the students' academic goals have a medium–large effect size (d = 0.36, rank 45). The students' other goal orientations (i.e., performance, learning, and avoidance goals), intrinsic academic motivation, and control expectations have small effects only (with d′s between 0.22 and 0.32). Extrinsic motivation and pessimistic attribution style are independent of achievement. Overall, these findings stress the importance of control cognitions (i.e., self-efficacy) and students' commitment to academic achievement in higher education.

Personality. Of the 11 categories in our review, personality was the category with the most meta-analytic findings and original studies. This category contains 16 variables and 694 effect sizes, based on a total of over one million students (see Table 3). However, only 25% of the variables have a medium–strong effect, and none has a strong effect.

The two variables with the largest absolute effect sizes in this category are conscientiousness (d = 0.47, rank 30) and test anxiety (d = −0.43, rank 104), followed by emotional intelligence and the need for cognition, which have about equally strong effect sizes (d = 0.35, rank 47). However, the confidence interval of need for cognition ranges from 0.05 to 0.66 leaving open whether this variable has a negligible, weak, or medium strong effect in the population. Conscientiousness is the tendency to be organized, achievement-focused, disciplined, and industrious. A study comparing various facets of conscientiousness found that specifically industriousness, control, cautiousness, tidiness, task planning, and perseverance were related to academic achievement (MacCann, Duckworth, & Roberts, 2009). Several of these facets also predicted variables such as students' class absences and disciplinary infractions, in addition to achievement. Among college students, conscientiousness is positively related to goal setting, that is, commitment to one's goals and orientation toward learning goals (Klein & Lee, 2006). In higher education, the association between conscientiousness and academic achievement is about as strong as the relation between intelligence and academic achievement (d = 0.47, rank 30, for both; see Table 1).

Test anxiety has a medium–strong negative relation with achievement. Test anxiety prevents students from applying their knowledge to exam tasks and thus leads to an underestimation of the students' true competence (for an overview see Zeidner, 1998). A meta-analysis of 77 studies with a total of 2,482 students (Ergene, 2003) shows that 4- to 6-hr psychological programs can significantly reduce test anxiety (d = −0.68, CI [−0.59, −0.77]). The programs that target students' concrete test-taking skills, in combination with their maladaptive cognition of the test situation have the strongest negative effect on anxiety (d = −1.22).

Six other variables in the personality category each have a positive but a rather weak correlation with achievement (i.e., locus of control, optimism, self-esteem, openness, agreeableness, and gender). Females have, on average, slightly higher achievement than males (d = 0.21, rank 76). Five personality variables are virtually independent of achievement (i.e., general self-concept, emotional stability, extraversion, depression, and biological age). To conclude, compared with other categories of student-related variables personality variables show rather weak relations with academic achievement. Exceptions are conscientiousness, test anxiety, emotional intelligence, and need for cognition, for which medium-sized effects were found.

Context variables. The context variables have the smallest effect sizes of all the 11 categories in Table 3. None of the 13 variables under this category has a strong effect. Only one variable has a medium–large effect size, specifically, to what extent a student received financial resources from an institution (d = 0.41, rank 37). This variable is confounded with the students' aptitude.

So it is unclear whether the effect size merely indicates a spurious correlation or whether financial support actually has a positive influence on achievement.

Discussion

Comparing Effect Sizes: A Cautionary Note

This article provides the first systematic international review of the meta-analyses on the variables associated with achievement in higher education (cf. Polanin et al., 2016). The 105 effect sizes in our main table of results allow for 5,460 pairwise comparisons of effect sizes, not all of which we can discuss here. Many readers will try to draw their own conclusions from these findings. However, the comparison of two ranks or of two effect sizes is not straightforward and needs to account for a number of questions (Coe, 2002; Ferguson, 2009).

Eight examples of the many questions that can be asked for any comparison of two effect sizes are: First, did the two meta-analyses include comparable control conditions or, for instance, did one compare against regular instruction and the other against no instruction? Second, was the treatment intensity (e.g., intervention duration) comparable in the two meta-analyses? Third, were the inclusion and exclusion criteria similar in the two meta-analysis? Fourth, can the difference in Cohen's *d* between two meta-analyses be attributed to random sampling error and measurement error (as indicated by the degree of overlap of the confidence intervals) or is the descriptive difference also of statistical significance? Fifth, is a statistically significant difference between two effect sizes sufficiently large to be practically relevant?

Sixth, is a difference in Cohen's *d* between two meta-analyses due to differences in the means or due to differences in the standard deviations of the measured variable? For instance, two meta-analyses finding the same absolute increase of achievement can still differ in their meta-analytically obtained values of Cohen's *d* when they focus on (sub)populations of students that differ in the variability of their achievement scores.

Seventh, do the two investigated constructs (e.g., teaching methods) have similar implementation costs? For example, beginning a lesson with an advance organizer only has an effect of $d = 0.26$ (Luiten et al., 1980), but costs little time and almost no other resources. In contrast, playing instructional games with virtual reality is more effective ($d = 0.51$), but has immense costs in terms of software development, computer hardware, and instructional time (Merchant et al., 2014). From a practical point of view, a less effective method with lower implementation costs can be preferable over a more effective method with higher implementation costs.

Finally, was a teaching method predominantly evaluated with specific content and might be less effective for other content? For example, a meta-analysis on games with virtual reality found a medium–large effect size (Merchant et al., 2014). However, all of the included original empirical studies used carefully selected learning goals which lend themselves ideally for gamification. The high effect size does not imply that replacing any part of any class with a virtual-reality game would increase student achievement (Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013). Moderator analyses showed that virtual reality games in combination with standard instruction are more effective than stand-alone virtual reality games. Thus, it is even possible that the gamification of a whole class would decrease student achievement.

Due to this inherent complexity of effect size comparisons, we urge practitioners and policymakers who try to draw their own inferences from our results to do so in collaboration with trained and experienced researchers from education, instructional psychology, or related fields.

Ten Cornerstone Findings

In the following, we discuss central findings of our study. We deliberately limit ourselves to covering 10 cornerstone findings, which have become apparent from the comparison of the 105 effect sizes and concern central topics in the learning sciences or ongoing debates in the research on higher education. By limiting the discussion to 10 key points, we try to avoid redundancies with the 38 included meta-analyses and to acknowledge the fact that it is impossible to present a comprehensive discussion of all possible 5,460 pairwise comparisons of our 105 variables.

1. There is broad empirical evidence related to the question what makes higher education effective. As can be observed from literature databases, the older publications about higher education mainly used theoretical approaches and neglected empirical research. This has thoroughly changed. In our systematic review, we synthesize 3,330 effect sizes from quantitative empirical studies involving a total of 1,920,239 students. Teachers in higher education can make their courses more effective by using these findings in the preparation and delivery of their courses. The methodological quality of the available studies and meta-analyses is mixed and needs to be considered in the interpretation of the empirical findings.

2. Most teaching practices have positive effect sizes, but some have much larger effect sizes than others. Of the 105 investigated variables, 87% have at least a small effect size (see Table 3). Thus, most teaching practices and learner characteristics are systematically related to achievement. Hattie (2009) already made a similar observation for education in general. As Hattie warned, teachers using almost any instructional method or tool might observe an impact on achievement and conclude that their instruction has an optimal effect. This could lead to the view that any teaching approach is acceptable. This perspective is wrong because some effect sizes are much larger than others. The real question is not whether an instructional method has an effect on achievement but whether it has a higher effect size than alternative approaches. Reviews of meta-analyses provide this information.

3. The effectivity of courses is strongly related to what teachers do. Previous reviews of the meta-analyses on the correlates of achievement, which mainly focused on school learning, found that proximal variables, that describe what teachers and students do and think during a lesson, are more closely related to achievement than distal variables, such as school demographics or state policies (Hattie, 2009, p. 22; Kulik & Kulik, 1989; Wang et al., 1993b, p. 74). Our results for higher education indirectly support this conclusion. Many variables in the categories social interaction, stimulating meaningful learning, and presentation relate to specific forms of teacher behavior, for example, asking open-ended questions instead of close-ended ones or writing a few keywords instead of half sentences on presentation slides. This fine-grained level of detail is sometimes called the microstructure of instruction. To be effective, teachers need to pay attention to the microstructure of their courses (Dumont & Istance, 2010).

Many of the included meta-analyses that explicitly report causal evidence from randomized controlled trials investigated elements of the microstructure of instruction, because these elements are easy to manipulate in experimental designs. Evidence on causal relations was reported for the teacher relating content to students, mastery learning, concept maps, intelligent tutoring systems, frequent testing, spoken explanation of visualizations, and animations (see Table 1).

As described in the introduction, it has been asked whether the choice of teaching methods is as important in higher education as it is in K–12 school education. After all, teachers and students are highly select groups with above average aptitude and skills. Our review indicates that even in higher education the choice of teaching methods has substantial effects on achievement.

4. The effectivity of teaching methods depends on how they are implemented. It is not only what teachers do on the microlevel but also how exactly they do it that critically affects achievement. Of all the meta-analyses in our review, 95% found at least one significant moderator effect. For example, when asking questions it is more effective to ask open-ended than close-ended questions (Redfield & Rousseau, 1981). Presentation slides with a few written keywords are more effective than presentation slides with written half sentences or full sentences (Adesope & Nesbit, 2012). Concept maps are more effective when they only depict central ideas and give no details (Nesbit & Adesope, 2006). Student projects that are prepared carefully and supervised closely by a teacher are much more effective than student projects requiring pure discovery learning (Alfieri et al., 2011). Giving an oral rather than a written explanation of a picture is more effective for dynamic visualizations (e.g., a film) than for static visualizations (e.g., a photograph; Reinwein, 2012). Many meta-analyses identified more than one moderator of the effectivity of an instructional method. For example, a meta-analysis on problem-based learning reports 48 effect sizes for the levels of various moderator variables (Dochy et al., 2003).

This shows that the same instructional method can have stronger or weaker effects on achievement, depending on how it is implemented on the microlevel. Meta-analytical results on the average effects of a teaching method, such as the ones provided in Table 1, do not suffice for designing effective classes. Educators should consider the moderator effects, and ultimately, they need the practical skills required for implementing effective methods in effective ways.

Fortunately, training programs for teachers in higher education can have broad beneficial effects on their classroom behavior, skills, and attitudes, as concluded in a literature review of 36 empirical studies (Stes, Min-Leliveld, Gijbels, & van Petegem, 2010). Training methods such as microteaching and similar approaches, where teachers receive both video and colleague feedback on the details of an instructional unit, have particularly strong positive effects inside (Brinko, 1993; Johannes & Seidel, 2012; Remesh, 2013) and outside (Fukkink, Trienekens, & Kramer, 2011; Hattie, 2009) higher education. Averaging over four meta-analyses with a total of 439 effect sizes, Hattie (2009, pp. 112–113) reports an effect of microteaching methods for K–12 school teachers on student achievement of $d = 0.88$, which was the fourth highest effect size in Hattie's list of 138 correlates of achievement.

If teachers' behaviors on the microlevel affect students' learning outcomes, then the time and effort that teachers invest in planning and organizing the microstructure of their courses should be strongly associated with student achievement. This is indeed the case (Feldman, 1989). Of our 105 variables, the teacher's preparation/organization of a course is the variable with the third largest effect size. The two variables with larger effects are correlates of achievement that cannot be directly influenced by a teacher. Thus, of all the variables in Table 1 that teachers can directly influence, the time and effort that they invest in the preparation of the microstructure of a course have the strongest effect. This is no surprise, as the design principles of effective instruction need to be carefully adapted to specific courses, content and student populations. This requires not only teachers' knowledge, skills, and creativity, but also their time and effort.

Feldman (1989) only presented very weak meta-analytical evidence from correlations, for example, without checking for publication bias. Nonetheless, it is still highly likely that the time and effort that a teacher invests in preparing the microstructure of instruction causally affect student achievement because a causal link between the microstructure and achievement has been firmly established in the literature (see cornerstone finding 3). As a consequence, higher education systems do not only have to give teachers knowledge of effective teaching methods and the skills for implementing them, but also time and incentives for planning and preparing the details of their courses.

5. Teachers can improve the instructional quality of their courses by making a number of small changes. The first four points imply that teachers can increase the instructional quality of their courses and their students' achievement by following a number of empirically supported design principles. Improving instructional quality in higher education does not always require revolutionary changes of the higher education system. Relatively small and easy to implement principles that each teacher can follow are:

- encouraging frequent class attendance (rank 6),
- stimulating questions and discussion (rank 11),
- providing clear learning goals and course objectives (rank 13),
- asking open-ended questions (rank 16),
- relating the content to the students (rank 17),
- providing detailed task-focused and improvement-oriented feedback (rank 30; see also Kluger & DeNisi, 1996),
- being friendly and respecting the students (rank 30),
- complementing spoken words with visualizations or written words (e.g., handouts or presentation slides; rank 42),
- using a few written keywords instead of half or full sentences on presentation slides (moderator analysis in Adesope & Nesbit, 2012),
- letting the students construct and discuss concept maps of central ideas covered in a course (rank 45),
- beginning each instructional unit with an advance organizer (rank 64), and
- avoiding distracting or seductive details in presentations (rank 101 and $d = -0.30$ for the presentation of seductive details).

Each of these points is relatively easy to implement. Our results indicate that many teachers using these behaviors in many of their classes might have a huge combined effect on student achievement in higher education.

6. The combination of teacher-centered and student-centered instructional elements is more effective than either form of instruction alone. No meta-analysis directly compared the effectivity of teacher-centered and student-centered instructional methods. However, a number of meta-analyses reported effect sizes for specific teacher-centered instructional elements, for example, teacher presentations, and student-centered instructional elements, for example, student projects. Six of the nine variables under the presentation category (teacher's clarity and understandableness; teacher's stimulation of interest in the course and its subject matter; teacher's elocutionary skills; teacher's enthusiasm for subject or teaching; spoken explanation of visualizations; and animations) had medium–large or large effect sizes (see Table 3). At the same time, all the five variables under the social interaction category (teacher's encouragement of questions and discussion; teacher's availability and helpfulness; open-ended questions; small-group learning; teacher's concern and respect for students, friendliness) had a medium–large or large effect size. Thus, both forms of instruction are effective.

Additional studies show that combinations of the two forms are even more effective than either form in isolation. For example, interactive elements increase the effectivity of lecture classes (Campbell & Mayer, 2009; Hake, 1998; King, 1990), and teacher-directed elements enhance the effectivity of student projects (Alfieri et al., 2011; Mayer, 2004). At least some of these studies (e.g., Campbell & Mayer, 2009) had experimental designs and, thus, indicate causal effects rather than mere correlations.

The effectivity of combining teacher-centered and student-centered instructional elements is also demonstrated by the high effect sizes found for the category social interaction. All five variables under this category have medium-large or large effect sizes, making it the category of instruction-related variables with the highest number of strong effects. Many meta-analyses and reviews on social forms of learning conclude that it is maximally effective when the teachers careful prepare and guide their students' activities and interactions while at the same time giving their students enough freedom to develop their own ideas and to make their own experiences. Thus, the combination of teacher-centered and student-centered elements is a necessary condition for social interaction to be effective (Hmelo-Silver, Duncan, & Chinn, 2007; Johnson, Johnson, & Smith, 2014; Slavin, 2010; Springer et al., 1999). From this perspective, lectures are still a timely and effective form of instruction provided they are given in an engaging and interactive way. Small-group learning and project-based learning are likewise timely and effective provided they are well prepared and closely supervised by the teacher.

7. Educational technology is most effective when it complements classroom interaction. Overall, the empirical results show that expanding the use of information and communication technology at the expense of other forms of instruction is likely to have detrimental effects on achievement. Of the six instruction-related variable categories, technology had the second lowest number of medium-large or large effect sizes. The categories social interaction, stimulating meaningful learning, assessment, and presentation had larger effect sizes. At the same time, the three variables with the highest effect sizes in the category technology have high implementation costs and are useful only for very specific content (games with virtual reality, rank 27; simulations with virtual reality, rank 37; intelligent tutoring systems, rank 47). Two concrete examples of the dangers of replacing classroom instruction with educational technology are that human tutors are more effective than intelligent tutoring software (moderator analysis in Steenbergen-Hu & Cooper, 2014) and that face-to-face small-group learning in the classroom (rank 27) is associated with higher achievement than small-group learning with technology (rank 77).

For online and classroom instruction, the empirical evidence indicates that they are most effective when they are combined into blended learning arrangements. Online learning (rank 88) is about as effective as classroom learning (Means et al., 2013), but blended learning (rank 52), that is, the combination of both forms of learning, is more effective than classroom instruction alone (Bernard et al., 2014). Thus, from the perspective of instructional effectiveness, the most important question for further research is not whether one form of instruction should replace the other but in which ways the two forms should be combined. As there are many forms of blended learning that differ in their educational approaches, complexity, and opportunities for social interaction (Bonk & Graham, 2006; Means et al., 2013) further meta-analyses with more detailed moderator analyses are needed.

The results presented here are representative only for instructional technology in the past, not for instructional technology in the present or future. Instructional technology advances quickly. The meta-analyses in this category had been conducted between 2001 and 2014. They could only include the studies that had already been published and thus were older than the meta-analyses themselves. Particularly, MOOCs, clickers, and social media have only been investigated in very few randomized controlled trials, if any, and no meta-analysis has been published yet. No definite evidence-based conclusions about their effectivity can be drawn yet (Hew & Cheung, 2013; Lantz, 2010; McAndrew & Scanlon, 2013).

Against this background it is interesting to test whether the effect sizes in the field of instructional technology increased over time due to the technological progress in the past. Four of the six meta-analyses in the technology category tested whether the effect sizes were related to the publication year. None of the four meta-analyses found an increase in effect sizes. The effect sizes significantly decreased from 1990 to 2011 for intelligence tutoring systems (Steenbergen-Hu & Cooper, 2014) and stayed constant from 1964 to 1999 for small-group learning with technology (Lou et al., 2001), from 1996 to 2008 for online learning (Means et al., 2013), and from 1990 to 2010 for blended learning (Bernard et al., 2014). Obviously, technological advances do not automatically improve educational practices or raise student achievement.

The empirical evidence in the technology category was of a good quality. Two of the meta-analyses in the technology category followed an above-average number of methodological recommendations and none followed a below-average number of recommendations (cf. Table 2). Our results also mirror the meta-analytic findings for instruction and communication technology in K–12 schools, that are characterized by small to medium effect sizes (Hattie, 2009, p. 201) and a lack of effect size increases over time (Hattie, 2009, p. 221). So far, there is no empirical support for the claim that instruction and communication technology will revolutionize higher education, at least not with respect to student achievement. Technology cannot be used to compensate for a lack of teachers or a lack of teacher training in higher education (see cornerstone findings 3 and 4). However, technology use can have weak to medium–strong associations with achievement when teachers use it in goal-directed ways as part of a carefully prepared overarching didactic concept.

8. Assessment practices are about as important as presentation practices. Our results indicate that assessment practices are related to achievement about as strongly as presentation practices. Thus, teachers should invest as much time in their assessment practices as they do in their presentations, which is currently not always the case. Assessment practices include not only giving exams but also setting explicit learning goals, establishing clear standards for success, and giving learning-oriented feedback (Boud & Falchikov, 2007). These instructional elements guide students' selective attention, the goals they set for themselves, and the way they prepare for exams (Broekkamp & van Hout-Wolters, 2007; Lundeberg & Fox, 1991). Thus, assessment practices do not only influence what students learn but also how they learn it (cf. cornerstone finding 10). Whereas the meta-analytic evidence in the area of higher education is of a mixed quality (see Table 2), the importance of assessment and feedback practices as determinants of students learning strategies, motivation, and achievement is well investigated for learning outside higher education and yielded comparable results (Clark, 2012; Hattie & Timperley, 2007; Kluger & DeNisi, 1996).

9. Intelligence and prior achievement are closely related to achievement in higher education. Among the five student-related variable categories, the category intelligence and prior achievement has the highest number of large effect sizes. That is, intelligence and prior achievement are of high predictive value and diagnostic importance when students are asking whether they should enroll in a higher education program. High school GPA and prior knowledge documented in admission test results show large effect sizes whereas general cognitive ability tests have a medium large effect size. It has to be taken into account that the effect size for cognitive ability tests has not been corrected for range restrictions (Poropat, 2009). Increasing levels of range restriction partly explain why effect sizes for the relation of cognitive ability tests with academic achievement decrease with increasing educational level (Jensen, 1998), for example, from $d = 1.42$ in primary education to $d = 0.49$ in secondary education and $d = 0.47$ in tertiary education in the study by Poropat (2009). While admission tests are specifically designed for student populations applying for higher education, cognitive ability tests are usually designed for broader populations, therefore suffering more strongly from range restrictions when applied to students in higher education. Further, with increasing educational level prior knowledge seems to gain importance (e.g., Dochy, de Rijdt, & Dyck, 2002) because tertiary education, more so than primary or secondary education, aims at equipping students with advanced knowledge in a specific content domain. Prior achievement or knowledge as well as intelligence are affected by the amount and quality of prior schooling (Ceci & Williams, 1997). Regarding achievement in higher education, colleges and universities therefore benefit from high instructional quality in K–12 schools and should try to improve this quality, for example, by establishing or contributing to excellent training programs for K–12 school teachers.

10. Students' strategies are more directly associated with achievement than students' personality or personal context. Students' strategies for learning and exam preparation, as well as for effort regulation and goal setting, tend to show stronger relations with achievement than their personalities or personal backgrounds, such as their gender and age. Students can change their strategies more easily than their personalities or personal backgrounds. Improving students' strategies in their regular classes in the context of their scientific discipline is more effective than training them in extracurricular settings with artificially created problems (Hattie et al., 1996; Tricot & Sweller, 2014). Among the most productive student strategies are frequent class attendance (Credé et al., 2010), effort regulation, a strategic approach to learning, and effective time and study management (Richardson et al., 2012). The two meta-analyses of Credé et al. (2010) and Richardson et al. (2012) were of a high methodological quality (cf. Table 2). Their findings indicate that working as hard as possible all of the time is not the best student strategy for high achievement. Instead, it is important to choose deliberately when and where to invest time and mental resources. Regarding student personality, conscientiousness shows the closest positive relation with academic achievement in higher education. The strength of this association remains fairly constant across educational levels, while associations between other personality traits within the Big Five model and academic performance decrease with increasing educational level (Poropat, 2009; see Walton & Billera, 2016, for a review).

Implications for Further Research

In recent years, several reviews of meta-analyses have been published inside (e.g., Hattie, 2009; Kulik & Kulik, 1989; Wang et al., 1993b) and outside (e.g., Delgado-Rodriguez, 2006; Hillberg, Hamilton-Giachritsis, & Dixon, 2011; Hyde, 2005) the field of education. By combining the results of several meta-analyses, each of which synthesizes several empirical studies, reviews of meta-analyses can give an overview of the number, range, stability, breadth, and strength of the effects in a research field. Reviews of meta-analyses in general and our study in particular also have some limitations.

First, more randomized controlled experiments are needed so that hypotheses about causal relations can be tested. Many meta-analyses did not present separate effect sizes for studies with experimental designs and with correlational designs. Only for 12 out of the 105 investigated variables do the meta-analyses report evidence of causal relations. There are several reasons for this. According to some authors, they could not analyze the results of controlled experiments with randomization, simply because these experiments were rare in their research area (Ruiz-Primo et al., 2011). Other meta-analyses focused on field studies—where randomization of the participants was not feasible—to maximize the ecological validity of their findings (Dochy et al., 2003). Several meta-analyses investigated stable personal characteristics that could not be manipulated experimentally, such as gender or personality traits (Poropat, 2009).

A second limitation, not only of our study but of educational studies in general, is that it is not easily possible to separate novelty effects from persistent effects. For example, a teacher who used to give 90-min lectures without any interactive elements suddenly delivers a lecture with a 15-min group discussion, which leads to higher learning gains. In this case, it is unclear whether group discussions in lectures are always effective, or whether they only captured the students' attention due to the teacher's attempt at something new. Generally, this problem leads to an overestimation of the effectiveness of new forms of instruction compared with more traditional forms.

A third drawback concerns the representativeness of our findings for current higher education worldwide. Our review could only include the meta-analyses that had been published in the past, which in turn covered only the empirical studies that had been published prior to that point. Although 23 of the 38 included meta-analyses had been published over the last 10 years, the timeliness of at least some of our findings might not be optimal, particularly the results about the educational technologies, that improve rapidly. Likewise, most of the empirical studies had been conducted in North America, fewer in Europe and Australia, and even scarcer in other continents, limiting our findings' generalizability to higher education worldwide.

Finally, all methodological limitations of meta-analyses also apply to our review of the meta-analyses. Not all the meta-analyses included in our review adhered to the established methodological standards (e.g., APA, 2008; Moher et al., 2009; Shea et al., 2009). On the positive side, the majority of the meta-analyses included grey literature to minimize the publication bias and reported explicit inclusion and exclusion criteria, confidence intervals around effect sizes, heterogeneity indices, and moderator analyses. On the negative side, less than half of the included meta-analyses indicated the exact search string they used, the interrater agreement for the coding of the effect sizes, the effect sizes corrected for the nonperfect reliabilities of the measures, an outlier or sensitivity analysis, or whether a random-effects model or a fixed-effects model had been used. Thus, it is important for future research on achievement in higher education to follow high-quality standards for research methods and reports of the results (cf. Ruiz-Primo et al., 2011). This will increase the replicability of the studies and facilitate future integrations of empirical findings (Open Science Collaboration, 2015).

Further typical shortcomings of meta-analyses and strategies for minimizing these problems are described by Ferguson and Brannick (2012), as well as Hunter and Schmidt (2004). A frequent critique about meta-analyses is that by averaging results over different samples and measures, they compare apples and oranges. Defendants of meta-analyses routinely reply that a researcher can only draw conclusions about the general characteristics of fruits by doing just that—comparing different types of fruits.

Conclusion

This study is the first systematic and comprehensive review of the meta-analyses on the variables associated with achievement in higher education published in the international research literature. The results are generally compatible with those of similar reviews for school learning or for learning in general. They offer evidence-based arguments for current debates in higher education. Instructional methods and the way they are implemented on the microlevel are substantially associated with achievement in higher education. This emphasizes the importance of teacher training in higher education. Among the different approaches to teaching, social interaction has the highest frequency of high positive effect sizes. Lectures, small-group learning, and project-based learning all have positive associations with achievement provided they balance teacher-centered with student-centered instructional elements. As yet, instruction and communication technology has comparably weak effect sizes, which did not increase over the past decades. High-achieving students in higher education are characterized by qualities that, in part, are affected by prior school education, for example, prior achievement, self-efficacy, intelligence, and the goal-directed use of learning strategies. Thus, universities indirectly benefit from a high instructional quality in K–12 schools. The application of our findings into practice is a complex process that requires the collaboration of practitioners, educational researchers, and policymakers. The effect sizes of the meta-analyses included in this systematic review indicate that such an evidence-based approach has great potential for increasing achievement in higher education.

References

Adesope, O. O., & Nesbit, C. (2012). Verbal redundancy in multimedia learning environments: A meta-analysis. *Journal of Educational Psychology, 104,* 250–263. http://dx.doi.org/10.1037/a0026147

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103,* 1–18. http://dx.doi.org/10.1037/a0021017

Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers*. San Francisco, CA: Jossey-Bass.

APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *American Psychologist, 63,* 839–851. http://dx.doi.org/10.1037/0003-066X.63.9.839

Bandura, A. (1979). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28,* 117–148. http://dx.doi.org/10.1207/s15326985ep2802_3

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C.-L. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research, 85,* 89–99. http://dx.doi.org/10.1080/00220671.1991.10702818

Barber, M., Donnelly, K., & Rizvi, S. (2013). *An avalanche is coming: Higher education and the revolution ahead*. London, UK: Institute for Public Policy Research.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods, 10,* 389–396. http://dx.doi.org/10.1037/1082-989X.10.4.389

Benassi, V. A., & Buskist, W. (2012). Preparing the new professoriate to teach. In W. A. Buskist & V. A. Benassi (Eds.), *Effective college and university teaching: Strategies and tactics for the new professoriate* (pp. 1–8). Los Angeles, CA: Sage. http://dx.doi.org/10.4135/9781452244006.n1

Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education, 26,* 87–122. http://dx.doi.org/10.1007/s12528-013-9077-3

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Maidenhead, UK: SRHE and Open University Press.

Bligh, D. A. (2000). *What's the use of lectures?* San Francisco, CA: Jossey-Bass.

Bonk, C. J., & Graham, C. R. (2006). *The Handbook of blended learning: Global perspectives, local designs*. San Francisco, CA: Pfeiffer.

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *Handbook of research synthesis and meta-analysis* (pp. 221–235). New York, NY: Russel Sage.

Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Behavior Research Methods, 8,* 5–18. http://dx.doi.org/10.1002/jrsm.1230

Boud, D., & Falchikov, N. (Eds.). (2007). *Rethinking assessment in higher education: Learning for the longer term*. London, UK: Routledge.

Bourhis, J., & Allen, M. (1992). Meta-analysis of the relationship between communication apprehension and cognitive performance. *Communication Education, 41,* 68–76. http://dx.doi.org/10.1080/03634529209378871

Brinko, K. T. (1993). The practice of giving feedback to improve teaching: What is effective? *The Journal of Higher Education, 64,* 574–593. http://dx.doi.org/10.2307/2959994

Broekkamp, H., & van Hout-Wolters, B. H. A. M. (2007). Students' adaptation of study strategies when preparing for classroom tests. *Educational Psychology Review, 19,* 401–428. http://dx.doi.org/10.1007/s10648-006-9025-0

Campbell, J., & Mayer, R. E. (2009). Questioning as an instructional method: Does it affect learning from lectures? *Applied Cognitive Psychology, 23,* 747–759. http://dx.doi.org/10.1002/acp.1513

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence and income. *American Psychologist, 52,* 1051–1058. http://dx.doi.org/10.1037/0003-066X.52.10.1051

Cherney, I. D. (2008). The effects of active learning on students' memories for course content. *Active Learning in Higher Education, 9,* 152–171. http://dx.doi.org/10.1177/1469787408090841

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1,* 73–105. http://dx.doi.org/10.1111/j.1756-8765.2008.01005.x

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24,* 205–249. http://dx.doi.org/10.1007/s10648-011-9191-6

Clark, J. (2008). PowerPoint and pedagogy: Maintaining student interest in university lectures. *College Teaching, 56,* 39–45. http://dx.doi.org/10.3200/CTCH.56.1.39-46

Clark, R. E. (1994). Media will never influence learning. *Educational Technology Research and Development, 42,* 21–29. http://dx.doi.org/10.1007/BF02299088

Coe, R. (2002, September). *It's the effect size, stupid*. Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, Exeter, UK. Retrieved from http://www.leeds.ac.uk/educol/documents/00002182.htm

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155

Cooper, H., & Koenka, A. C. (2012). The overview of reviews: Unique challenges and opportunities when research syntheses are the principal elements of new integrative scholarship. *American Psychologist, 67,* 446–462. http://dx.doi.org/10.1037/a0027119

Credé, M., & Kuncel, N. R. (2008). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science, 3,* 425–453. http://dx.doi.org/10.1111/j.1745-6924.2008.00089.x

Credé, M., Roch, S. G., & Kieszczynka, U. M. (2010). Class attendance in college: A meta-analytic review of the relationship of class attendance with grades and student characteristics. *Review of Educational Research, 80,* 272–295. http://dx.doi.org/10.3102/0034654310362998

Delgado-Rodríguez, M. (2006). Systematic reviews of meta-analyses: Applications and limitations. *Journal of Epidemiology and Community Health, 60,* 90–92. http://dx.doi.org/10.1136/jech.2005.035253

Dochy, F., de Rijdt, C., & Dyck, W. (2002). Cognitive prerequisites and learning: How far have we progressed since Bloom? Implications for educational practice and teaching. *Active Learning in Higher Education, 3,* 265–284. http://dx.doi.org/10.1177/1469787402003003006

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning and Instruction, 13,* 533–568. http://dx.doi.org/10.1016/S0959-4752(02)00025-7

Dumont, H., & Istance, D. (2010). Analysing and designing learning environments for the 21st century. In H. Dumont, D. Istance, & F. Benavides (Eds.), *The nature of learning: Using research to inspire practice* (pp. 19–34). Paris, France: OECD. http://dx.doi.org/10.1787/9789264086487-3-en

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56,* 455–463. http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x

Ergene, T. (2003). Effective interventions on test anxiety reduction: A meta-analysis. *School Psychology International, 24,* 313–328. http://dx.doi.org/10.1177/01430343030243004

European Commission. (2008). *The European Qualifications Framework for Lifelong Learning (EQF)*. Luxembourg, Europe: Office for Official Publications of the European Communities.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research, 59,* 395–430. http://dx.doi.org/10.3102/00346543059004395

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70,* 287–322. http://dx.doi.org/10.3102/00346543070003287

Feldman, K. A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education, 30,* 583–645. http://dx.doi.org/10.1007/BF00992392

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40,* 532–538. http://dx.doi.org/10.1037/a0015808

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods, 17,* 120–128. http://dx.doi.org/10.1037/a0024445

Fukkink, R. G., Trienekens, N., & Kramer, L. J. C. (2011). Video feedback in education and training: Putting learning in the picture. *Educational Psychology Review, 23,* 45–63. http://dx.doi.org/10.1007/s10648-010-9144-5

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15,* 313–331. http://dx.doi.org/10.1016/j.learninstruc.2005.07.001

Gottfried, A. E., Fleming, J. S., & Gottfried, A. W. (2001). Continuity of academic intrinsic motivation from childhood through late adolescence: A longitudinal study. *Journal of Educational Psychology, 93,* 3–13. http://dx.doi.org/10.1037/0022-0663.93.1.3

Groot, W., & Maassen van den Brink, H. (2007). The health effects of education. *Economics of Education Review, 26,* 186–200. http://dx.doi.org/10.1016/j.econedurev.2005.09.002

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66,* 64–74. http://dx.doi.org/10.1119/1.18809

Hambrick, D. Z. (2003). Why are some people more knowledgeable than others? A longitudinal study of knowledge acquisition. *Memory & Cognition, 31,* 902–917. http://dx.doi.org/10.3758/BF03196444

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.

Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66,* 99–136. http://dx.doi.org/10.3102/00346543066002099

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77,* 81–112. http://dx.doi.org/10.3102/003465430298487

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92,* 490–499. http://dx.doi.org/10.1037/0033-2909.92.2.490

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504. http://dx.doi.org/10.1037/1082-989X.3.4.486

Helle, L., Tynjälä, P., & Olkinuora, E. (2006). Project-based learning in post-secondary education: Theory, practice and rubber sling shots. *Higher Education, 51,* 287–314. http://dx.doi.org/10.1007/s10734-004-6386-5

Hew, K. F., & Cheung, W. S. (2013). Use of Web 2.0 technologies in K–12 and higher education: The search for evidence-based practice. *Educational Research Review, 9,* 47–64. http://dx.doi.org/10.1016/j.edurev.2012.08.001

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327,* 557–560. http://dx.doi.org/10.1136/bmj.327.7414.557

Hillberg, T., Hamilton-Giachritsis, C., & Dixon, L. (2011). Review of meta-analyses on the association between child sexual abuse and adult mental health difficulties: A systematic approach. *Trauma, Violence & Abuse, 12,* 38–49. http://dx.doi.org/10.1177/1524838010386812

Hmelo-Silver, C. E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16,* 235–266. http://dx.doi.org/10.1023/B:EDPR.0000034022.16470.f3

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42,* 99–107. http://dx.doi.org/10.1080/00461520701263368

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction, 17,* 722–738. http://dx.doi.org/10.1016/j.learninstruc.2007.09.013

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks, CA: Sage. http://dx.doi.org/10.4135/9781412985031

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60,* 581–592. http://dx.doi.org/10.1037/0003-066X.60.6.581

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Johannes, C., & Seidel, T. (2012). Professionalisierung von Hochschullehrenden: Lehrbezogene Vorstellungen, Wissensanwendung und Identitätsentwicklung in einem videobasierten Qualifikationsprogramm [Professionalization of university teachers: Teaching approaches, professional vision and teacher identity in video-based training]. *Zeitschrift für Erziehungswissenschaft, 15,* 233–251. http://dx.doi.org/10.1007/s11618-012-0273-0

Johnson, D. W., Johnson, R. T., & Smith, K. A. (2014). Cooperative learning: Improving university instruction by basing practice on validated theory. *Journal on Excellence in College Teaching, 25,* 85–118.

Kalechstein, A. D., & Nowicki, S., Jr. (1997). A meta-analytic examination of the relationship between control expectancies and academic achievement: An 11-year follow-up to Findley and Cooper. *Genetic, Social, and General Psychology Monographs, 123,* 27–56.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19,* 528–558. http://dx.doi.org/10.1080/09541440601056620

King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal, 27,* 664–687. http://dx.doi.org/10.3102/00028312027004664

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal, 29,* 303–323. http://dx.doi.org/10.3102/00028312029002303

Kirschner, P. A., & Paas, F. (2001). Web-enhanced higher education: A tower of Babel. *Computers in Human Behavior, 17,* 347–353. http://dx.doi.org/10.1016/S0747-5632(01)00009-7

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41,* 75–86. http://dx.doi.org/10.1207/s15326985ep4102_1

Klein, H. J., & Lee, S. (2006). The effects of personality on learning: The mediating role of goal setting. *Human Performance, 19,* 43–66. http://dx.doi.org/10.1207/s15327043hup1901_3

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119,* 254–284. http://dx.doi.org/10.1037/0033-2909.119.2.254

Kobayashi, K. (2005). What limits the encoding effect of note-taking? A meta-analytic examination. *Contemporary Educational Psychology, 30,* 242–262. http://dx.doi.org/10.1016/j.cedpsych.2004.10.001

Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60,* 265–299. http://dx.doi.org/10.3102/00346543060002265

Kulik, J. A., & Kulik, C.-L. C. (1989). The concept of meta-analysis. *International Journal of Educational Research, 13,* 221–340. http://dx.doi.org/10.1016/0883-0355(89)90052-9

Kuncel, N. R., Kochevar, R. J., & Ones, D. S. (2014). A meta-analysis of letters of recommendation in college and graduate admissions: Reasons for hope. *International Journal of Selection and Assessment, 22,* 101–107. http://dx.doi.org/10.1111/ijsa.12060

Lamb, A. C. (2006). Multimedia and the teaching-learning process in higher education. *New Directions for Teaching and Learning, 51,* 33–42.

Lantz, M. E. (2010). The use of 'Clickers' in the classroom: Teaching innovation or merely an amusing novelty? *Computers in Human Behavior, 26,* 556–561. http://dx.doi.org/10.1016/j.chb.2010.02.014

Larwin, K. H., Gorman, J., & Larwin, D. A. (2013). Assessing the impact of testing aids on post-secondary student performance: A meta-analytic investigation. *Educational Psychology Review, 25,* 429–443. http://dx.doi.org/10.1007/s10648-013-9227-1

Levasseur, D. G., & Sawyer, J. K. (2006). Pedagogy meets PowerPoint: A research review of the effects of computer-generated slides in the classroom. *Review of Communication, 6,* 101–123. http://dx.doi.org/10.1080/15358590600763383

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71,* 449–521. http://dx.doi.org/10.3102/00346543071003449

Luiten, J., Ames, W., & Ackerson, G. (1980). A meta-analysis of the effects of advance organizers on learning and retention. *American Educational Research Journal, 17,* 211–218. http://dx.doi.org/10.3102/00028312017002211

Lundeberg, M. A., & Fox, P. W. (1991). Do laboratory findings on test expectancy generalize to classroom outcomes? *Review of Educational Research, 61,* 94–106. http://dx.doi.org/10.3102/00346543061001094

MacCann, C., Duckworth, A. L., & Roberts, R. D. (2009). Empirical identification of the major facets of conscientiousness. *Learning and*

Individual Differences, 19, 451–458. http://dx.doi.org/10.1016/j.lindif.2009.03.007

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *The British Journal of Educational Psychology, 81,* 59–77. http://dx.doi.org/10.1348/000709910X503501

Marton, F., & Säljö, R. (1976). On qualitative differences in learning – 1: Outcome and process. *The British Journal of Educational Psychology, 46,* 4–11. http://dx.doi.org/10.1111/j.2044-8279.1976.tb02980.x

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59,* 14–19. http://dx.doi.org/10.1037/0003-066X.59.1.14

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38,* 43–52. http://dx.doi.org/10.1207/S15326985EP3801_6

McAndrew, P., & Scanlon, E. (2013). Education. Open learning at a distance: Lessons for struggling MOOCs. *Science, 342,* 1450–1451. http://dx.doi.org/10.1126/science.1239686

Means, B., Toyama, Y., Murphy, R., & Baki, M. (2013). The effectiveness of online and blended learning: A meta-analysis of the empirical literature. *Teachers College Record, 15,* 1–47.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K–12 and higher education: A meta-analysis. *Computers & Education, 70,* 29–40. http://dx.doi.org/10.1016/j.compedu.2013.07.033

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151,* 264–269, W64. http://dx.doi.org/10.7326/0003-4819-151-4-200908180-00135

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement, 56,* 63–75. http://dx.doi.org/10.1177/0013164496056001004

Neisser, U., Boodoo, G., Jr., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychological Association, 51,* 77–101. http://dx.doi.org/10.1037/0003-066X.51.2.77

Nesbit, J. C., & Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76,* 413–448. http://dx.doi.org/10.3102/00346543076003413

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31,* 199–218. http://dx.doi.org/10.1080/03075070600572090

Organization for Economic Co-operation and Development (OECD). (2012). *Education indicators in focus*. Retrieved from http://www.oecd.org/edu/skills-beyond-school/educationindicatorsinfocus.htm

Organization for Economic Co-operation and Development (OECD). (2014). *Education at a glance 2014: OECD indicators*. Paris, France: OECD Publishing.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349,* no pagination. http://dx.doi.org/10.1126/science.aac4716

Pappano, L. (2012, November 2). *Year of the MOOC. New York Times*. Retrieved from http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html

Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36,* 89–101. http://dx.doi.org/10.1207/S15326985EP3602_4

Perry, R. P., & Smart, J. C. (Eds.). (2007). *The scholarship of teaching and learning in higher education: An evidence-based perspective*. Dordrecht, the Netherlands: Springer. http://dx.doi.org/10.1007/1-4020-5742-3

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16,* 385–407. http://dx.doi.org/10.1007/s10648-004-0006-x

Pintrich, P. R., & Zusho, A. (2007). Student motivation and self-regulated learning in the college classroom. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 731–810). New York, NY: Springer. http://dx.doi.org/10.1007/1-4020-5742-3_16

Poitras, J. (2012). Meta-analysis of the impact of the research setting on conflict studies. *International Journal of Conflict Management, 23,* 116–132. http://dx.doi.org/10.1108/10444061211218249

Polanin, J. R., Maynard, B. R., & Dell, N. A. (2016). Overviews in educational research: A systematic review and analysis. *Review of Educational Research*. Advance online publication. http://dx.doi.org/10.3102/0034654316631117

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135,* 322–338. http://dx.doi.org/10.1037/a0014996

Pressley, M., Harris, K. R., & Marks, M. B. (1992). But good strategy instructors are constructivists! *Educational Psychology Review, 4,* 3–31. http://dx.doi.org/10.1007/BF01322393

Redfield, D. L., & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research, 51,* 237–245. http://dx.doi.org/10.3102/00346543051002237

Reinwein, J. (2012). Does the modality effect exist? And if so, which modality effect? *Journal of Psycholinguistic Research, 41,* 1–32. http://dx.doi.org/10.1007/s10936-011-9180-4

Remesh, A. (2013). Microteaching, an efficient technique for learning effective teaching. *Journal of Research in Medical Sciences, 18,* 158–163.

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review, 7,* 216–237. http://dx.doi.org/10.1016/j.edurev.2012.05.003

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138,* 353–387. http://dx.doi.org/10.1037/a0026838

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin, 130,* 261–288. http://dx.doi.org/10.1037/0033-2909.130.2.261

Robbins, S. B., Oh, I.-S., Le, H., & Button, C. (2009). Intervention effects on college performance and retention as mediated by motivational, emotional, and social control factors: Integrated meta-analytic path analyses. *Journal of Applied Psychology, 94,* 1163–1184. http://dx.doi.org/10.1037/a0015738

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52,* 59–82. http://dx.doi.org/10.1146/annurev.psych.52.1.59

Rücker, G., Schwarzer, G., Carpenter, J. R., & Schumacher, M. (2008). Undue reliance on I-squared in assessing heterogeneity may mislead. *BMC Medical Research Methodology*. Advance online publication.

Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impact of undergraduate science course innovations on learning. *Science, 331,* 1269–1270. http://dx.doi.org/10.1126/science.1198976

Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin, 135,* 1–22. http://dx.doi.org/10.1037/a0013978

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Thousand Oaks, CA: Sage.

Schneider, M. (2012). Knowledge integration. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 1684–1686). New York, NY: Springer.

Schneider, M., & Mustafić, M. (2015). *Gute Hochschullehre: Eine evidenzbasierte Orientierungshilfe* [Effective teaching in higher education: An evidence-based perspective]. Heidelberg, Germany: Springer.

Schünemann, H. J., Oxman, A. D., Vist, G. E., Higgins, J. P. T., Deeks, J. J., Glasziou, P., & Guyatt, G. H. (2011). Interpreting results and drawing conclusions. In J. P. T. Higgins & S. Green (Eds.), *Cochrane handbook for systematic reviews of interventions (Version 5.1.0)*. London, UK: Cochrane Collaboration.

Schwartz, B. M., & Gurung, R. A. R. (Eds.). (2012). *Evidence-based teaching for higher education*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/13745-000

Schwinger, M., Wirthwein, L., Lemmer, G., & Steinmayr, R. (2014). Academic self-handicapping and achievement: A meta-analysis. *Journal of Educational Psychology, 106,* 744–761. http://dx.doi.org/10.1037/a0035832

Seidel, T., Rimmele, R., & Prenzel, M. (2005). Clarity and coherence of lesson goals as a scaffold for student learning. *Learning and Instruction, 15,* 539–556. http://dx.doi.org/10.1016/j.learninstruc.2005.08.004

Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., . . . Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology, 62,* 1013–1020. http://dx.doi.org/10.1016/j.jclinepi.2008.10.009

Slavin, R. E. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin, 94,* 429–445. http://dx.doi.org/10.1037/0033-2909.94.3.429

Slavin, R. E. (2010). Co-operative learning: What makes group-work work?. In H. Dumont, D. Istance, & F. Benavides (Eds.), *The nature of learning: Using research to inspire practice* (pp. 161–178). Paris, France: OECD Publishing. http://dx.doi.org/10.1787/9789264086487-9-en

Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69,* 21–51. http://dx.doi.org/10.3102/00346543069001021

Steenbergen, B., Crajé, C., Nilsen, D. M., & Gordon, A. M. (2009). Motor imagery training in hemiplegic cerebral palsy: A potentially useful therapeutic tool for rehabilitation. *Developmental Medicine and Child Neurology, 51,* 690–696. http://dx.doi.org/10.1111/j.1469-8749.2009.03371.x

Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology, 106,* 331–347. http://dx.doi.org/10.1037/a0034752

Steinmayr, R., Meißner, A., Weidinger, A. F., & Wirthwein, L. (2014). Academic achievement. *Oxford Bibliographies*. Retrieved from http://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0108.xml

Stes, A., Min-Leliveld, M., Gijbels, D., & van Petegem, P. (2010). The impact of instructional development in higher education: The state-of-the-art of the research. *Educational Research Review, 5,* 25–49. http://dx.doi.org/10.1016/j.edurev.2009.07.001

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin, 121,* 371–394. http://dx.doi.org/10.1037/0033-2909.121.3.371

Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review, 26,* 265–283. http://dx.doi.org/10.1007/s10648-013-9243-1

UNESCO. (2012). *International standard classification of education (ISCED) 2011*. Montreal, Canada: UNESCO Institute for Statistics.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140,* 1174–1204. http://dx.doi.org/10.1037/a0036620

Wagner, E., & Szamosközi, Ş. (2012). Effects of direct academic motivation-enhancing intervention programs: A meta-analysis. *Journal of Cognitive and Behavioral Psychotherapies, 12,* 1584–1701.

Walton, K. E., & Billera, K. A. (2016). Personality development during the school-aged years: Implications for theory, research, and practice. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Eds.), *Psychosocial skills and school systems in the 21st century. Theory, research, and practice* (pp. 93–111). New York, NY: Springer. http://dx.doi.org/10.1007/978-3-319-28606-8_4

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993a). Toward a knowledge base for school learning. *Review of Educational Research, 63,* 249–294. http://dx.doi.org/10.3102/00346543063003249

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993b). What helps students learn? *Educational Leadership, 51,* 74–79.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105,* 249–265. http://dx.doi.org/10.1037/a0031311

Zeidner, M. (1998). *Test anxiety: The state of the art*. New York, NY: Plenum Press.

Zimmerman, B. J., & Schunk, D. H. (Eds.). (2011). *Handbook of self-regulation of learning and performance*. New York, NY: Routledge.

Received August 31, 2015
Revision received December 1, 2016
Accepted December 20, 2016 ■