Review of Educational Research
Month 201X, Vol. XX, No. X, pp. 1–38
DOI: 10.3102/0034654315605917

Does College Teach Critical Thinking? A Meta-Analysis

Christopher R. Huber and Nathan R. Kuncel
University of Minnesota

Educators view critical thinking as an essential skill, yet it remains unclear how effectively it is being taught in college. This meta-analysis synthesizes research on gains in critical thinking skills and attitudinal dispositions over various time frames in college. The results suggest that both critical thinking skills and dispositions improve substantially over a normal college experience. Furthermore, analysis of curriculum-wide efforts to improve critical thinking indicates that they do not necessarily produce incremental long-term gains. We discuss implications for the future of critical thinking in education.

Keywords: critical thinking, college students, changes in critical thinking

Educators, policymakers, and employers have demonstrated a sustained interest in teaching critical thinking, as both an important life skill and an asset to the future workforce (Koenig et al., 2011). This interest is particularly evident in college, where critical thinking has gained traction as a crucial component of general education (Arum & Roksa, 2011; Halpern, 2001). A recent study reported that faculty endorsed teaching critical thinking as the most important goal of undergraduate education, with over 99% describing it as "very important" or "essential" (DeAngelo et al., 2009, p. 3). Critical thinking is also viewed as an important component of many medium- and high-complexity jobs (Peterson et al., 1997). However, despite the value placed on teaching critical thinking, the actual effectiveness of college at doing so—through either explicit instruction or general exposure—remains a disputed point.

Numerous interventions have been tested to increase both critical thinking skills and the general disposition toward critical thinking. McMillan (1987) reviewed the literature and concluded that specific interventions to foster critical thinking were not well supported but college attendance itself did appear to have a bolstering effect. Later research has suggested that some college experiences may actually affect gains on critical thinking. Terenzini, Springer, Pascarella, and

Nora (1995) reported that both out-of-class experiences, such as number of unassigned books read, and in-class variables, such as type of courses taken, explained significant variance in critical thinking ability after a year of college. In a meta-analysis, Abrami et al. (2008) also found a modest but positive effect for critical thinking interventions. In addition, the researchers found a large amount of variability in effect sizes, much of which was attributable to the type of intervention and the degree of implementation. This finding suggests that larger gains on critical thinking might be achieved by focusing on the most effective types of intervention, should current gains be deemed inadequate. Finally, in an unpublished meta-analysis, Ortiz (2007) found larger mean effect sizes in samples where a critical thinking intervention was employed.

Even without explicit attempts to foster critical thinking, there is certainly a widespread perception that college breeds critical thinkers. Tsui (1998) reported that 92% of students in a large multi-institution study believed they had made some gains in critical thinking, and 39.3% thought their critical thinking had grown much stronger. Only 8.9% believed it had not changed or had grown weaker. However, it is not clear whether students share a common definition of critical thinking and whether they are capable of an accurate self-assessment. The actual effects of college on global critical thinking remain unclear. First, there is disagreement about the magnitude of critical thinking gains over the course of college. Second, there is disagreement about whether the concept of global critical thinking makes sense in the first place. This study synthesizes effect sizes to estimate the magnitude of gains on general critical thinking measures and provides a theoretical basis for interpreting these results.

Changes in Critical Thinking During College

Past quantitative reviews support the hypothesis that college improves critical thinking. Ortiz's (2007) meta-analysis estimates gains of 0.12 standard deviations (*SD*s) per semester for nonphilosophy students. This finding supports the general efficacy of college for teaching critical thinking. One concern, however, is that studies that occurred over different intervals of time were simply rescaled to a single semester effect. It is unclear whether critical thinking increases linearly over the college years or more over certain intervals. Arum and Roksa (2011) suggested that it may increase more in the beginning, but other evidence is mixed (Pascarella & Terenzini, 1991). Given that the rate of change may differ depending on the time frame studied, analyses should ideally avoid collapsing data from different time frames into a single global estimate.

The problem with collapsing across time frames is further illustrated by reversing Ortiz's (2007) procedure and estimating the effects over 4 years of college. If an effect size measured over 4 years can be divided by 8 and combined with semester-long studies, then the reverse must also hold. However, rescaling Ortiz's semester effect sizes to 4-year effects leads to estimates that are implausibly large. For example, Ortiz reported semester-long gains of 0.78 *SD*s in philosophy courses that incorporate large amounts of argument mapping practice. Expanded to a full 4-year education, this would imply that consistent argument mapping practice could result in massive gains of 6.24 *SD*s in critical thinking skill. This thought experiment does not imply that large changes could not occur over a single semester. Rather, it suggests that large gains due to short-term interventions may not be sustained or expanded on indefinitely. In other words, there is probably some upper limit to the amount of critical thinking skills that will be learned and retained in college, even if critical thinking instruction is explicitly infused into the curriculum.

Other research suggests that legitimate gains do occur even without explicit critical thinking instruction. Pascarella and Terenzini (1991, 2005) reviewed evidence from several studies, which on the whole indicate that college improves critical thinking. However, Pascarella and Terenzini (2005) noted that the gains observed in their review of 1990s studies are "appreciably smaller in magnitude than the gains we observed in our previous synthesis" (p. 158). Specifically, they estimated that seniors had an advantage of 1 *SD* in critical thinking ability over freshmen in the pre-1990s samples compared to around half an *SD* in the 1990s.[1]

Arum and Roksa (2011) cited this decline as possible evidence that college is actually becoming *less* effective at teaching critical thinking over time. Their own large-scale study reveals gains of 0.18 *SD*s on the Collegiate Learning Assessment (CLA) performance task over the first three semesters of college and 0.47 *SD*s over 4 years, which they deem insufficient. However, the CLA is a measure of skills such as written communication and complex reasoning in addition to critical thinking (Arum & Roksa, 2011; Klein, Benjamin, Shavelson, & Bolus, 2007). Such criterion contamination makes it impossible to directly compare Arum and Roksa's (2011) results to those of Pascarella and Terenzini (1991, 2005) or to interpret them purely as indicators of critical thinking gains. Nevertheless, Arum and Roksa's estimates are in keeping with the half *SD* estimate from Pascarella and Terenzini (2005; Arum & Roksa, 2011, Footnote 14, p. 219).

In light of Arum and Roksa's (2011) dissatisfaction with observed gains on the CLA, it is helpful to consider what magnitude of gains might be reasonably expected. One way to do this is to examine the effects of other variables in college. A second-order meta-analysis by Hattie (1992) found that the average effect of numerous situational and individual difference variables on academic achievement was 0.40 *SD*s (see also Hattie, 1987). Even students' general disposition toward learning, arguably a central predictor of success, had an overall effect size of only 0.61. In this context, we have little reason to expect gains larger than half an *SD* on critical thinking. Even with explicit instruction, producing meaningfully larger gains might be difficult.

Changes in Critical Thinking Across Majors

Some have suggested that certain majors may produce larger gains in critical thinking than others. If this were the case, an analysis of majors that produced the strongest gains would be useful. First, it would suggest the magnitude of gains that could be achieved with the correct curriculum (although it may be possible to improve upon even the best current curriculum). Second, understanding the features that distinguish gain-producing majors would inform future attempts to improve critical thinking in other majors.

The current evidence on differences between majors is inconclusive. Pascarella and Terenzini's (2005) review failed to find strong evidence for differential gains across majors. By contrast, Ortiz's (2007) meta-analysis suggests that philosophy students may learn more critical thinking than other students. Ortiz estimated gains of 0.26 *SD*s per semester for philosophy students compared to only 0.12 *SD*s per semester for other majors. However, the number of pure philosophy samples in her analysis is small (k = 6), and the samples appear to be from unpublished studies. In addition, the confidence intervals for the philosophy and nonphilosophy effect sizes show substantial overlap. Ortiz herself suggested that the observed difference may simply be statistical noise.

Another likely exemplar of critical thinking instruction is nursing programs. The National League for Nursing (NLN) requires that nursing programs include formal critical thinking training (Adams, Whitlow, Stover, & Johnson, 1996). This requirement makes nursing students a useful comparison group for estimating the long-term effects of sustained instruction in critical thinking. Measuring this long-term change has different implications than measuring the effects of short-term interventions (e.g., adding a critical thinking component to a single course). For example, short-term critical thinking instruction may give students an initial advantage that does not ultimately persist after the posttest, which could occur either because the benefits are temporary or because other students eventually catch up. Such a result would be analogous to recent findings from the Head Start impact study. Puma et al. (2012) found that the Head Start program yielded substantial advantages in preschool children, but most of the benefits dissipated by the third grade. Similarly, it is possible that students in a critical thinking–rich curriculum may enjoy an initial head start that is negated in the long run.

Although this possibility would ideally be tested by a true experimental design, such studies are rare (if not nonexistent) in this literature due to the difficulty of enacting an institution-wide curricular experiment. However, a comparison between nursing students and students in other majors is a feasible proxy for such a study. With their strong emphasis on critical thinking, nursing programs can be seen as representative of a college education in which explicit critical thinking instruction is the norm rather than the subject of an occasional intervention. We would therefore expect greater gains on critical thinking in nursing programs to the extent that formal training is incrementally effective. Currently, the literature is lacking any comprehensive comparison of such programs to a more traditional college experience.

Definition and Measurement Issues

Another difficulty in the critical thinking literature is defining the construct. The traditional generalist view conceptualizes critical thinking as a broad ability to interpret information and approach problems correctly that can be applied across a wide variety of domains (e.g., McMillan, 1987; Pascarella, 1989). For example, Abrami et al. (2008) defined critical thinking as "the ability to engage in purposeful, self-regulatory judgment" (p. 1102). Researchers have distinguished between critical thinking skills and dispositions, suggesting a meaningful distinction between the ability to think critically and willingness to actually do so. Most measures focus on the skill aspect, with the California Critical Thinking Disposition Inventory (CCTDI) being the main exception (N. C. Facione, Facione, & Sanchez, 1994).

Some scholarship has questioned traditional conceptualizations of critical thinking as a broad domain-general skill. McPeck (1984, 1990) argued that critical thinking has been operationally reduced to the ability to analyze arguments. According to this perspective, the ability to reason and think critically is required for a broad range of tasks beyond analysis of logical arguments, such as "finding one's way home, investing money, fishing, driving a car, doing sums, shopping, playing hopscotch, intelligent voting, building math models, writing poems, and countless other classes of activities" (McPeck, 1984, p. 30). McPeck argued that the ability to think critically about such a broad array of domains is not well represented by any general skill (e.g., analyzing arguments), and therefore critical thinking ability is best conceptualized as domain-specific.

Kuncel (2011) argued that when people describe critical thinking skills, they refer to one of two very different things. The first is field- or job-specific expertise, which will form only with practice and experience in a field and is not widely generalizable. For example, thinking critically about medical diagnosis in veterinary medicine is hard-earned expertise that does not readily transfer to restaurant management. Kuncel suggested that critical thinking in college should emphasize field-specific expertise at least as much as the second type of critical thinking. The second type, as currently measured by many critical thinking scales, is "a finite set of very specific reasoning skills (e.g., gambler's fallacy, law of large numbers, correlation vs. causation)" (Kuncel, 2011, p. 2). Although these skills are viewed as useful, it is argued that knowing about the law of large numbers is useful for reasoning about the law of large numbers and nothing else. This narrow definition also calls into question the degree to which critical thinking can be taught in a broad way that will transfer to improved performance across all work, school, or life tasks. In his meta-analytic review, Kuncel found little discriminant validity evidence for commonly used critical thinking tests and other cognitive ability assessments. In addition, the evidence suggested that these tests (and gains on them) are unlikely to predict grades or job performance more effectively than common measures of IQ or general cognitive ability (although they may be useful for specific tasks).

Despite questions about the scope of critical thinking skills, most researchers have argued that critical thinking tests do measure useful traits (e.g., P. A. Facione, 2011). Even if critical thinking skills are domain-specific, the specific reasoning skills measured by commonly used tests are likely to produce more informed consumers of information (Kuncel, 2011). In addition, the attitudinal disposition toward critical thinking is more likely to apply across domains than specific critical thinking skills. If college can promote general skepticism toward questionable claims and ideas, especially ones that mesh with one's own worldview, it has surely performed a valuable function. Although individuals lack the specific knowledge needed to critically analyze every domain, a disposition toward critical thinking should at least encourage acquisition of additional knowledge and reservation of judgment.

Although there is little disagreement that critical thinking is important, teaching it takes time away from teaching other important skills, such as reading and mathematics. Given these trade-offs, it is important to understand the present state of critical thinking in college and what can reasonably be done if it is inadequate.

To do this, we must determine whether a normal college education is even effective at teaching critical thinking. In addition, we must estimate the incremental gains over longer periods of time when more resources are devoted to critical thinking instruction. This meta-analytic study will establish average gains on tests of general critical thinking during college, both with and without formal training, and reconcile conflicting views in this domain.

Method

Search Strategy and Inclusion Criteria

Our search included studies that reported critical thinking skills or dispositions at multiple points during a student's tenure in college. Searches were conducted using PsycInfo (1872–2012), ERIC (1966–2012), ProQuest Dissertations and Theses (1861–2012), and Google Scholar (through 2012). To maximize the number of relevant studies found, the broad search terms "college students" and "critical thinking" were used on the first three databases. Due to the large number of irrelevant results, "critical thinking" was revised to "changes in critical thinking" for the Google Scholar search. We also included papers found in the literature reviews and reference sections of other articles.

Each study was coded for means and *SD*s of critical thinking scores across different points of time in college. All longitudinal and cross-sectional designs were included that provided sufficient information to calculate a standardized mean difference between at least two points in time or class years (e.g., comparing current freshmen to current seniors). Measures of domain-specific critical thinking were not included. Examples of excluded measures include tests of critical thinking in psychology and critical thinking essays in English for foreign-language speakers (Wang & Liao, 2012). We also excluded one time point from Bartlett and Cox (2002) because it represented gains after clinical rather than academic experiences.

Studies that presented only partial change data were also excluded from our sample. For example, Mentkowski and Rogers (1985) reported gains on three of the five Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1980) subscales but did not include data for the other two subscales or overall score gains. The purpose of this exclusion was to avoid upwardly biasing the effect size estimates by considering only the statistically significant changes from a given study. Studies that relied only on self-reported changes in critical thinking and those that only reported results from experimental interventions to increase critical thinking were also excluded from the meta-analytic sample. However, control groups from experimental studies were included, as long as no explicit effort was made to teach critical thinking skills or dispositions to these control groups.

The one exception to the exclusion of explicit interventions was studies of baccalaureate nursing students. Although individual courses specifically geared toward teaching critical thinking to nursing students were still excluded, the NLN specifically includes critical thinking instruction in its accreditation requirements. The high interest in teaching critical thinking among nursing programs is reflected in the relatively large number of studies that used nursing samples. As such, nursing students could be analyzed separately to address possible concerns about contamination (i.e., gains due to critical thinking instruction

TABLE 1
Frequencies of exclusion criteria applied to prescreened studies

Exclusion criterion	No. of studies	% excluded studies
Does not measure critical thinking at multiple time points	27	30.7
Only reports gains after an intervention[a]	19	21.6
Noncollege sample	10	11.4
Missing vital information	10	11.4
No data (research proposals, qualitative studies)	6	6.8
Only uses a domain-specific critical thinking measure	6	6.8
Not related to critical thinking	4	4.5
Self-report data only	3	3.4
Time frame too short (less than one term)	2	2.3
Only reports statistically significant subscale gains	1	1.1
Total	88	100.0

[a]Includes logic/reasoning courses.

rather than the normal college experience) and as a useful comparison group to demonstrate the possibilities for explicitly teaching critical thinking over the span of a college education.

Ultimately, we excluded 88 studies that were deemed potentially eligible for inclusion after our initial review. Table 1 lists the frequencies of various rationales for rejecting studies. Although some studies may be deficient in multiple categories, we present the primary reason for rejection for each study. The most common exclusion factors were measuring critical thinking at only one time point (30.7%), only reporting gains in critical thinking after an intervention (21.6%), using a noncollege sample (11.4%), and missing other vital information needed to calculate an effect size (11.4%). After our exclusion criteria were applied, a total of 71 studies remained for the final analysis.

Effect Size Calculation

Cohen's d was calculated for changes in critical thinking across varying time frames. This effect size quantifies the difference between two means in *SD* units. However, standardized effect sizes can be calculated in different metrics depending on the available data and the specific research question. In this case, we considered two possibilities described by Morris and DeShon (2002): the raw score metric and the change score metric. The raw score d value is calculated by dividing the difference in mean scores between two time points by a pooled *SD*. Variations on this effect size are commonly used in critical thinking research. The change score d value is calculated using the *SD* of change scores in the denominator rather than an average *SD* of observed scores. Effect sizes in the change score metric tend to be larger and are appropriate for analyzing change over time (Morris & DeShon, 2002).

Although the *SD* of change scores is almost never reported in the critical thinking literature, it is possible to convert raw score effect sizes to change score effect sizes under certain conditions. Morris and DeShon (2002) presented a formula to perform the conversion when the correlation between pretest and posttest scores can be estimated in each study. A meta-analytic estimate of this correlation in longitudinal studies can be applied to perform the conversion for cross-sectional effect sizes. However, this method is not recommended if there is significant residual variance in pretest–posttest correlations across studies.

To investigate the possibility of using the change score metric, we performed a preliminary meta-analysis of pretest–posttest correlations from 29 longitudinal studies in our sample. The sampling variance of a correlation varies according to the population value of the correlation, particularly at values greater than .50 (Raudenbush & Bryk, 2002). Therefore, we used Fisher's r-to-z transformation to normalize the sampling variances prior to conducting the analysis. Since Fisher's transformed r approaches infinity as the raw correlation approaches one, two correlation estimates were reduced from 1 to .99 (the next highest value in our sample). We then conducted a random effects meta-analysis with time frame as a moderator and transformed the results back into the raw correlation metric.

These preliminary results did not support the use of the change score metric. Significant residual heterogeneity remained after accounting for the (nonsignificant) effect of time frame, $Q(27) = 1022.6326$, $p < .0001$. Given the residual variance in effect sizes across studies, it would be inappropriate to use the meta-analytic average as a proxy in studies without the data to compute a correlation directly. As such, we chose to use the more traditional raw score formula for Cohen's d.

In some cases, we still had to impute certain values necessary to estimate an effect size. When only a pretest *SD* was available, it was used as a stand-in for the posttest *SD*. When *SD*s were not available, estimates were taken from comparable samples with large sample sizes. Table 2 reports samples for which this method was used, as well as the source of imputed *SD*s. We also conducted a sensitivity analysis to determine if our results were robust to changes in the imputed values.

Once all effect sizes were calculated, we corrected the estimates for unreliability in the measures used. Measurement error systematically reduces effect size estimates, causing the magnitude of the true population effect to be underestimated (Hunter & Schmidt, 2004). This biasing effect can be corrected by dividing an effect size by the square root of the associated measure's reliability (Bobko, Roth, & Bobko, 2001). Whenever possible, we used study-specific internal consistency reliability estimates (i.e., coefficient alpha, Kuder-Richardson-20, Kuder-Richardson-21, and split-half) to correct for attenuation due to measurement error. When reliability statistics were not reported, we used estimates from test manuals. To avoid overcorrection, the most conservative (i.e., largest) estimates available were used in these cases. Four studies used less common measures of critical thinking skills for which reliability evidence was not available. We used a reliability estimate of .85 in these cases. This estimate was chosen because it was the most conservative imputed value among all other critical thinking skill measures.

TABLE 2

Source of imputed standard deviations by study

Study	Sample	Measure	Imputed *SD* source	Source sample
Berger (1984)	137 baccalaureate nursing students (sophomore-senior years)	Watson-Glaser	Watson and Glaser (1980)	266 baccalaureate nursing students
Dale, Ballotti, Handa, and Zych (1997)	39 college freshmen	Watson-Glaser	Watson and Glaser (1980)	824 freshmen in 4-year colleges
Drouin (1992), sopho-mores	120 second-semester college sophomores	Watson-Glaser	Watson and Glaser (1980)	824 freshmen in 4-year colleges
Drouin (1992), seniors	120 second-semester college seniors	Watson-Glaser	Watson and Glaser (1980)	417 upper-division students in 4-year colleges
Leppa (1997), CCTDI Group 1, Time 1	70 baccalaureate nursing students with Associate's degrees	CCTDI	N. C. Facione and Facione (1997)	817 juniors in nursing programs
Leppa (1997), CCTDI group 1, Time 2	70 baccalaureate nursing students with Associate's degrees	CCTDI	N. C. Facione and Facione (1997)	1,035 seniors in nursing programs
Leppa (1997), CCTDI Group 2, Time 1	77 baccalaureate nursing students with Associate's degrees	CCTDI	N. C. Facione and Facione (1997)	817 juniors in nursing programs
Leppa (1997), CCTDI Group 2, Time 2	77 baccalaureate nursing students with Associate's degrees	CCTDI	N. C. Facione and Facione (1997)	1,035 seniors in nursing programs
Leppa (1997), CCTST, Time 1	70 baccalaureate nursing students with Associate's degrees	CCTST	N. C. Facione and Facione (1997)	1,618 juniors in nursing programs
Leppa (1997), CCTST, Time 2	70 baccalaureate nursing students with Associate's degrees	CCTST	N. C. Facione and Facione (1997)	2,611 seniors in nursing programs
Pierce (1994), chemistry	33 students in a chemistry course	CCTST	Jacobs (1995)	1,383 entering college freshmen
Pierce (1994), psychology	17 students in a psychology course	CCTST	Jacobs (1995)	1,383 entering college freshmen

(continued)

TABLE 2 (continued)

Study	Sample	Measure	Imputed *SD* source	Source sample
Richards (1977)	72 nursing students (junior–senior year)	Watson-Glaser	Watson and Glaser (1980)	266 baccalaureate nursing students
Thompson and Rebeschi (1999), CCTDI, Time 1	38 nursing juniors	CCTDI	N. C. Facione and Facione (1997)	817 juniors in nursing programs
Thompson and Rebeschi (1999), CCTDI, Time 2	38 nursing seniors	CCTDI	N. C. Facione and Facione (1997)	1,035 seniors in nursing programs
Thompson and Rebeschi (1999), CCTST, Time 1	38 nursing juniors	CCTST	N. C. Facione and Facione (1997)	1,618 juniors in nursing programs
Thompson and Rebeschi (1999), CCTST, Time 2	38 nursing seniors	CCTST	N. C. Facione and Facione (1997)	2,611 seniors in nursing programs
Williams (2003), Sample 1 CCTST	138 students in a large educational psychology course	CCTST	Jacobs (1995)	1,383 entering college freshmen
Williams (2003), Sample 2 Watson-Glaser	149 students in a large educational psychology course	Watson-Glaser	Williams, Oliver, and Stockdale (2004)	69 psychology students[a]

Note. CCTDI = California Critical Thinking Disposition Inventory; CCTST = California Critical Thinking Skills Test.
[a]This source was chosen because the study was conducted by the same researcher as the study requiring imputation and used a similar sample.

Meta-Analysis

We conducted a mixed-effects multivariate meta-analysis using the METAFOR package in R (Viechtbauer, 2014). This method facilitates the aggregation of effect size measures from multiple primary studies and allows for multiple moderators to be analyzed in a single hierarchical model. An important feature of multivariate meta-analysis is that it permits analysis of multiple effect sizes drawn from overlapping samples. When more than two times or class years were included in a single study, we calculated *d*s between the initial time (e.g., freshman year) and each subsequent time separately. Because the resulting effect sizes share a common pretest group, their sampling errors are no longer independent (Gleser & Olkin, 2009). This violates the independence assumptions of most meta-analytic methods.

Multivariate meta-analysis incorporates dependent effect sizes by accounting for the covariances between their sampling errors. This is accomplished by constructing a variance-covariance matrix containing estimated sampling variances for each *d* value, as well as covariances between dependent effect size estimates. Our first step was to estimate sampling variances for each study based on sample size and study design. We estimated sampling variances using Morris and DeShon's (2002) Equation 22, which can be used for both cross-sectional and longitudinal designs. Sampling variance estimates for each effect size are included in Table 3.

Next, we estimated covariances between dependent effect sizes. Gleser and Olkin (2009) presented two relevant equations to estimate these covariances. First, Equation 19.19 estimates the covariance between effect sizes in multiple-treatment studies due to a shared control group. We used this equation to estimate covariance due to a common pretest group (or a common comparison group in cross-sectional studies). Second, Equation 19.27 can be used when multiple outcome measures are given to a single sample. In the present study, Mines et al. (1990) contributed data from two separate critical thinking measures given to a single sample. Estimated covariances between all dependent effect sizes are reported in Table 4.

As an additional check against homogeneous effect sizes, we specified a hierarchical data structure with effect sizes grouped within studies. Even if they are calculated from nonoverlapping samples, multiple effect size estimates from a single study may be more homogeneous than expected by chance. Using study of origin as a grouping variable accounts for increased similarity between effect sizes due to their common source.

Seventy-one studies met the criteria to be included in the meta-analysis (see Table 3). In total, we coded 124 effect sizes from these studies. Given evidence that critical thinking skills and dispositions are distinct constructs that behave differently with external correlates (Kuncel, 2011), we kept skills and dispositions separate for all analyses. We analyzed 110 effect sizes for critical thinking skills and 14 for dispositions.

Because our analysis mixes cross-sectional and longitudinal studies, the interpretation of a given study's sample size depends on the study design. For example, 100 participants in a cross-sectional study provide an average of 50 data

TABLE 3
Summary of studies in the meta-analysis

Study	Effect size (d)	σ_d^2	Pre n	Post n	Publication type	Grade	Time frame (Years)	Design	Nursing (yes/no)	Measure
Adams, Stover, and Whitlow (1999)	0	.00498	203	203	Journal	So–Sr	2.5	L	Y	Watson-Glaser
Arburn and Bethel (1999)	0.32	.03082	37	37	Conference paper	N/A	0.5	L	N	CCTST
Bartlett and Cox (2002), CCTDI	0.67	.04916	28	28	Journal	So	1	L	N	CCTDI
Bartlett and Cox (2002), CCTST	0.22	.04348	26	26	Journal	So	1	L	N	CCTST
Beckie, Lowry, and Barnett (2001), Cohort 1, 1 year	−0.09	.01891	55	55	Journal	Jr–Sr	1	L	Y	CCTST
Beckie et al. (2001), Cohort 1, 2 years	0.05	.01901	55	55	Journal	Jr–Sr	2	L	Y	CCTST
Beckie et al. (2001), Cohort 2, 1 year	0.07	.01896	55	55	Journal	Jr–Sr	1	L	Y	CCTST
Beckie et al. (2001), Cohort 2, 2 years	0.14	.01916	55	55	Journal	Jr–Sr	2	L	Y	CCTST
Beckie et al. (2001), Cohort 3, 1 year	−0.33	.01524	73	73	Journal	Jr–Sr	1	L	Y	CCTST
Beckie et al. (2001), Cohort 3, 2 years	−0.65	.01847	73	73	Journal	Jr–Sr	2	L	Y	CCTST
Berger (1984)	0.41	.00816	137	137	Journal	So–Sr	3	L	Y	Watson-Glaser
Blaich and Wise (2008)	0.11	.00033	3,081	3,081	Research report	Fr–So	1	L	N	CAAP-CT

(continued)

TABLE 3 (continued)

Study	Effect size (*d*)	σ_d^2	Pre *n*	Post *n*	Publication type	Grade	Time frame (Years)	Design	Nursing (yes/no)	Measure
Blaich and Wise (2011)	0.44	.00051	2,200	2,200	Research report	Fr–Sr	4	L	N	CAAP-CT
Brigham (1989), 1 year	0.6	.07842	28	28	Dissertation	Fr–So	1	C	Y	Watson-Glaser
Brigham (1989), 2 years	0.5	.07567	28	29	Dissertation	Fr–Jr	2	C	Y	Watson-Glaser
Brigham (1989), 3 years	0.74	.07907	28	29	Dissertation	Fr–Sr	3	C	Y	Watson-Glaser
Burgess (2003)	0.45	.02025	59	59	Dissertation	Jr–Sr	2	L	Y	Critical Thinking Assessment
Colbert (1987)	−0.41	.00768	146	146	Dissertation	N/A	0.5	L	N	Watson-Glaser
Colucciello (1997), CCTDI	0.18	.01750	129	105	Journal	Jr–Sr	1	C	Y	CCTDI
Colucciello (1997), CCTST	0.14	.01749	129	105	Journal	Jr–Sr	1	C	Y	CCTST
Criner (1992)	0.27	.05268	22	22	Dissertation	N/A	0.5	L	N	New Jersey Test of Reasoning Skills
Dale, Ballotti, Handa, and Zych (1997)	−0.08	.02718	39	39	Journal	Fr	0.5	L	N	Watson-Glaser
Daly (2001)	0.03	.02444	43	43	Journal	Fr–So	1.5	L	Y	Watson-Glaser
Drouin (1992)	0.2	.01691	120	120	Dissertation	So–Sr	2	C	N	Watson-Glaser
Erickson (1999)	0.09	.05621	20	20	Dissertation	Fr	0.5	L	N	Watson-Glaser
Ewen (2001), 1 year	−0.04	.02261	74	113	Dissertation	Fr–So	1	L	Y	Watson-Glaser
Ewen (2001), 2 years	−0.2	.02299	74	110	Dissertation	Fr–Jr	2	L	Y	Watson-Glaser
Ewen (2001), 3.5 years	−0.19	.02256	74	115	Dissertation	Fr–Sr	3.5	L	Y	Watson-Glaser
N. C. Facione and Facione (1997), CCTDI, 1 year	0.42	.00653	216	570	Book	Fr–So	1	C	Y	CCTDI

(continued)

TABLE 3 (continued)

Study	Effect size (d)	σ_d^2	Pre n	Post n	Publication type	Grade	Time frame (Years)	Design	Nursing (yes/no)	Measure
N. C. Facione and Facione (1997), CCTST, 1 year	0.3	.00393	311	1,485	Book	Fr–So	1	C	Y	CCTST
N. C. Facione and Facione (1997), CCTDI, 2 years	0.34	.00593	216	817	Book	Fr–Jr	2	C	Y	CCTDI
N. C. Facione and Facione (1997), CCTST, 2 years	0.44	.00391	311	1,618	Book	Fr–Jr	2	C	Y	CCTST
N. C. Facione and Facione (1997), CCTDI, 3 years	0.53	.00573	216	1,035	Book	Fr–Sr	3	C	Y	CCTDI
N. C. Facione and Facione (1997), CCTST, 3 years	0.54	.00367	311	2,611	Book	Fr–Sr	3	C	Y	CCTST
Fleeger (1986), 1 year	−0.26	.07082	41	23	Dissertation	Fr–So	1	C	Y	Watson-Glaser
Fleeger (1986), 2 years	0.25	.06395	41	27	Dissertation	Fr–Jr	2	C	Y	Watson-Glaser
Forbes (1997)	0.04	.05596	20	20	Dissertation	Fr	0.5	L	N	CCTST
Frost (1991), College 1	0.08	.01240	83	83	Journal	Fr	0.5	L	N	Watson-Glaser
Frost (1991), College 2	−0.05	.00864	118	118	Journal	Fr	0.5	L	N	Watson-Glaser
Giancarlo and Facione (2001)	0.28	.00720	147	147	Journal	Fr–Sr	3.5–4[a]	L	N	CCTDI
Goodin (2005)	0	.07212	16	16	Dissertation	Fr	0.5	L	Y	CCTST
Gross Takazawa, and Rose (1987), associate's degree	1.26	.05753	37	37	Journal	Fr–So	2	L	Y	Watson-Glaser
Gross et al. (1987), bachelor's degree	1.5	.07650	34	34	Journal	Jr–Sr	2	L	Y	Watson-Glaser

(continued)

TABLE 3 (continued)

Study	Effect size (d)	σ_d^2	Pre n	Post n	Publication type	Grade	Time frame (Years)	Design	Nursing (yes/no)	Measure
Hagedorn et al. (1999)	0.37	.00103	1,054	1,054	Journal	Fr–Jr	3	L	N	CAAP-CT
Hancock (1981)	−0.16	.00832	124	124	Dissertation	N/A	0.5	L	N	Watson-Glaser
Hill (1995)	0.69	.24968	10	10	Dissertation	Fr–Sr	4	C	N	Cornell Z
Isaacs (1990)	−0.07	.03596	30	30	Dissertation	N/A	0.5	L	Y	Cornell Z (translated into Spanish)
Johnson (2002)	0.48	.03892	99	37	Dissertation	All	1	L	Y	CCTST
Jones and Morris (2007)	−0.45	.01731	60	60	Journal	Fr–So	2	L	Y	Critical Thinking Assessment
Keller (1993)	−0.02	.02275	46	46	Dissertation	N/A	0.5	L	Y	Watson-Glaser
Kintgen-Andrews (1988), associate's degree in nursing	0.05	.01891	55	55	Research report	So	1	L	Y	Watson-Glaser
Kintgen-Andrews (1988), bachelor's degree in nursing	−0.06	.03724	29	29	Research report	So	1	L	Y	Watson-Glaser
Kintgen-Andrews (1988), practical nursing	0.07	.01894	55	55	Research report	Fr	1	L	Y	Watson-Glaser
Kintgen-Andrews (1988), prehealth science	0.06	.02788	38	38	Research report	Fr	1	L	N	Watson-Glaser
Kokinda (1989), 1 year	1.34	.31829	9	10	Dissertation	Fr–So	1	C	Y	Watson-Glaser
Kokinda (1989), 2 years	0.18	.19655	9	15	Dissertation	Fr–Jr	2	C	Y	Watson-Glaser
Kokinda (1989), 3 years	0.37	.20005	9	15	Dissertation	Fr–Sr	3	C	Y	Watson-Glaser

(continued)

TABLE 3 (continued)

Study	Effect size (d)	σ_d^2	Pre n	Post n	Publication type	Grade	Time frame (Years)	Design	Nursing (yes/no)	Measure
Lehmann (1963), females	0.95	.00333	461	461	Journal	Fr–Sr	4	L	N	Test of Critical Thinking
Lehmann (1963), males	1.07	.00286	590	590	Journal	Fr–Sr	4	L	N	Test of Critical Thinking
Leppa (1997), CCTDI Group 1	0.54	.01727	70	70	Journal	Jr–Sr	1	L	Y	CCTDI
Leppa (1997), CCTDI Group 2	0.66	.01675	77	77	Journal	Jr–Sr	1	L	Y	CCTDI
Leppa (1997), CCTST	−0.04	.01473	70	70	Journal	Jr–Sr	1	L	Y	CCTST
Lewis and Dahl (1972)	0.17	.00212	482	482	Journal	N/A	0.5	L	N	Watson-Glaser, Forms Ym and Zm
Lyons (2008)	0.42	.04574	27	27	Journal	So	0.5	L	Y	Critical Thinking Assessment
McCarthy, Schuster, Zehr, and McDougal (1999), CCTDI	0.34	.01860	156	85	Journal	So–Sr	3	C	Y	CCTDI
McCarthy et al. (1999), CCTST	0.54	.01919	156	85	Journal	So–Sr	3	C	Y	CCTST
McDonough (1997), nursing	2.02	.11345	30	30	Dissertation	N/A	2[b]	C	Y	Watson-Glaser
McDonough (1997), business	0.02	.06905	30	30	Dissertation	N/A	2[b]	C	N	Watson-Glaser
McDonough (1997), liberal arts	0.85	.07696	30	30	Dissertation	N/A	2[b]	C	N	Watson-Glaser
McDonough (1997), science	0.79	.07579	30	30	Dissertation	N/A	2[b]	C	N	Watson-Glaser

(continued)

TABLE 3 (continued)

Study	Effect size (d)	σ_d^2	Pre n	Post n	Publication type	Grade	Time frame (Years)	Design	Nursing (yes/no)	Measure
McKee (2006), Sample 1	0.13	.01384	75	75	Dissertation	N/A	0.5	L	N	Scottsdale Test of Critical Thinking
McKee (2006), Sample 2	0.14	.01101	94	94	Dissertation	N/A	0.5	L	N	Scottsdale Test of Critical Thinking
Meiss and Bates (1984)	−0.11	.03482	31	31	Conference paper	N/A	0.5	L	N	Watson-Glaser
Melvin (1996), Fr–So	0.1	.02491	92	73	Dissertation	Fr–So	2	C	N	Watson-Glaser
Melvin (1996), Fr–Jr	0.3	.02306	92	86	Dissertation	Fr–Jr	2	C	N	Watson-Glaser
Melvin (1996), Fr–Sr	0.58	.02791	92	65	Dissertation	Fr–Sr	4	C	N	Watson-Glaser
Miller (1987)	0.2	.00758	137	137	Dissertation	Fr–Sr	4	L	Y	Watson-Glaser
Mines, King, Hood, and Wood (1990), Cornell *Z*	1.76	.11867	20	40	Journal	Fr–Sr	3	C	N	Cornell Z
Mines et al. (1990), Watson-Glaser	2.22	.13337	20	40	Journal	Fr–Sr	3	C	N	Watson-Glaser
Nathan (1997), Sample 1	1.01	.11551	17	17	Dissertation	Sr	1	L	Y	Watson-Glaser
Nathan (1997), Sample 2	0.01	.05589	20	20	Dissertation	Jr	1	L	Y	Watson-Glaser
Norwood (1998)	0.15	.01662	63	63	Dissertation	So–Sr	2.5	L	Y	Watson-Glaser
Notarianni (1991)	−0.12	.00316	321	321	Dissertation	All	1	L	Y	Watson-Glaser
Pascarella (1989)	0.78	.09579	17	17	Journal	Fr	1	L	N	Watson-Glaser
Pierce (1994), chemistry	−0.18	.03313	33	33	Research report	N/A	0.5	L	N	CCTST
Pierce (1994), psychology	0.59	.08694	17	17	Research report	N/A	0.5	L	N	CCTST
Pierce (2011)	0.12	.00702	146	146	Dissertation	N/A	0.5	L	N	Cornell Z

(continued)

TABLE 3 (continued)

Study	Effect size (d)	σ_d^2	Pre n	Post n	Publication type	Grade	Time frame (Years)	Design	Nursing (yes/no)	Measure
Richards (1977)	−0.26	.01487	72	72	Journal	Jr–Sr	2	L	Y	Watson-Glaser
Rose (1997)	−0.86	.09586	18	18	Dissertation	Fr	0.5	L	N	Watson-Glaser
Ross and Semb (1981)	0.09	.02455	43	43	Journal	N/A	0.5	L	N	Watson-Glaser
Ruff (2005), CCTDI	−0.09	.05952	19	19	Dissertation	Fr	0.5	L	N	CCTDI
Ruff (2005), CCTST	−0.29	.06333	19	19	Dissertation	Fr	0.5	L	N	CCTST
Seifert, Pascarella, Colangelo, and Assouline (2007), honors	0.07	.00389	260	260	Journal	Fr	1	L	N	CAAP-CT
Seifert et al. (2007), nonhonors	0.05	.00058	1,739	1,739	Journal	Fr	1	L	N	CAAP-CT
Shin, Lee, Ha, and Kim (2006), 1 year	−0.24	.03460	32	32	Journal	Fr–So	1	L	Y	CCTDI
Shin et al. (2006), 2 years	−0.03	.03342	32	32	Journal	Fr–Jr	2	L	Y	CCTDI
Shin et al. (2006), 3 years	0.12	.03369	32	32	Journal	Fr–Sr	3	L	Y	CCTDI
Smith (1995)	0.01	.01739	140	99	Dissertation	Fr–So	1	C	Y	Watson-Glaser
Solon (2001)	0.08	.04203	26	26	Journal	N/A	0.5	L	N	Cornell Z
Solon (2003)	−0.18	.04293	26	26	Journal	N/A	0.5	L	N	Cornell Z
Solon (2007)	0.1	.04213	26	26	Journal	Fr–So	0.5	L	N	Cornell Z
Spaulding and Kleiner (1992), business	0.7	.16945	27	9	Journal	N/A	1.5[c]	C	N	Cornell Z[d]
Spaulding and Kleiner (1992), health science	0.61	.11379	38	13	Journal	N/A	1.5[c]	C	N	Cornell Z[d]

(continued)

TABLE 3 (continued)

Study	Effect size (d)	σ_d^2	Pre n	Post n	Publication type	Grade	Time frame (Years)	Design	Nursing (yes/no)	Measure
Spaulding and Kleiner (1992), liberal arts	−0.1	.17277	27	8	Journal	N/A	1.5[c]	C	N	Cornell Z[d]
Spaulding and Kleiner (1992), math/physical science	−0.43	.21978	15	8	Journal	N/A	1.5[c]	C	N	Cornell Z[d]
Spaulding and Kleiner (1992), social science	0.65	.10175	27	19	Journal	N/A	1.5[c]	C	N	Cornell Z[d]
Sukspringarm (1976)	0.02	.02783	38	38	Dissertation	Fr	0.5	L	N	Watson-Glaser
Sullivan (1987)	0	.02042	51	51	Journal	Jr–Sr	2	L	Y	Watson-Glaser
Thompson and Rebeschi (1999), CCTDI	0.33	.02967	38	38	Journal	Jr–Sr	1.5	L	Y	CCTDI
Thompson and Rebeschi (1999), CCTST	0.47	.03175	38	38	Journal	Jr–Sr	1.5	L	Y	CCTST
Tomlinson-Keasey and Eisert (1978), Sample 1	0.05	.02227	47	47	Research report	Fr	1	L	N	Watson-Glaser
Tomlinson-Keasey and Eisert (1978), Sample 2	0.02	.04565	24	24	Research report	Fr	1	L	N	Watson-Glaser
Turner (2000)	0.14	.03388	32	32	Journal	Fr–Sr	4	L	Y	CCTST
Watson and Glaser (1980)	−0.07	.01067	96	96	Test manual	N/A	0.5	L	N	Watson-Glaser
Wheeler and Collins (2003)	0.06	.03348	32	32	Journal	Jr	0.5	L	Y	CCTST
Whitmire (2001), 1.5 years	0.09	.00096	1,046	1,046	Journal	Fr–So	1.5	L	N	CAAP-CT
Whitmire (2001), 3.5 years	0.38	.00104	1,046	1,046	Journal	Fr–Sr	3.5	L	N	CAAP-CT

(continued)

TABLE 3 (continued)

Study	Effect size (d)	σ_d^2	Pre n	Post n	Publication type	Grade	Time frame (Years)	Design	Nursing (yes/no)	Measure
Williams (2003), Sample 1 CCTST	−0.31	.00788	138	138	Research report	N/A	0.5	L	N	CCTST
Williams (2003), Sample 2 Watson-Glaser	0.26	.00709	149	149	Research report	N/A	0.5	L	N	Watson-Glaser, Form S
Williams et al. (2004)	0.12	.01508	69	69	Journal	All	0.5	L	N	Watson-Glaser, Form S
Winter and McClelland (1978), community college	0.01	.06681	30	32	Journal	Fr–So	1	C	N	Test of Thematic Analysis
Winter and McClelland (1978), Ivy college	0.51	.01915	121	100	Journal	Fr–Sr	3	C	N	Test of Thematic Analysis
Winter and McClelland (1978), teachers' college	0.26	.06215	30	38	Journal	Fr–Sr	3	C	N	Test of Thematic Analysis
Yuan, Kunaviktikul, Klunklin, and Williams (2008)	−0.07	.04800	23	23	Journal	N/A	0.5	L	Y	CCTST

Note. C = cross-sectional; L = longitudinal; Fr = freshman; So = sophomore; Jr = junior; Sr = senior; N/A = not applicable; CCTST = California Critical Thinking Skills Test; CCTDI = California Critical Thinking Disposition Inventory; CAAP-CT = Collegiate Assessment of Academic Proficiency Critical Thinking test. Uncorrected effect sizes are reported here. Pre n = sample size at Time 1; post n = sample size at Time 2; σ_d^2 = estimated sampling variance of the effect size.

[a]Coded as 3.75 years. [b]Approximately 2 years. Comparison groups are students with fewer than 10 credits and more than 50 credits. [c]Approximately 1.5 years. Comparison groups are students with fewer than half of required core credit hours and students with more than half. [d]Due to limited testing time, an odd/even split-half was used so that each person got one half.

TABLE 4

Estimated sampling covariances between dependent effect sizes

Covarying effect sizes	σ_{d1d2}
Beckie, Lowry, and Barnett (2001), Cohort 1, 1 and 2 years	.01816
Beckie et al. (2001), Cohort 2, 1 and 2 years	.01823
Beckie et al. (2001), Cohort 3, 1 and 2 years	.01440
Blaich and Wise (2008); Blaich and Wise (2011)	.00046
Brigham (1989), 1 and 2 years	.03778
Brigham (1989), 1 and 3 years	.03878
Brigham (1989), 2 and 3 years	.03821
Ewen (2001), 1 and 2 years	.01353
Ewen (2001), 1 and 3.5 years	.01353
Ewen (2001), 2 and 3.5 years	.01359
N. C. Facione and Facione (1997), CCTDI, 1 and 2 years	.00468
N. C. Facione and Facione (1997), CCTDI, 1 and 3 years	.00470
N. C. Facione and Facione (1997), CCTDI, 2 and 3 years	.00468
N. C. Facione and Facione (1997), CCTST, 1 and 2 years	.00323
N. C. Facione and Facione (1997), CCTST, 1 and 3 years	.00324
N. C. Facione and Facione (1997), CCTST, 2 and 3 years	.00325
Fleeger (1986), 1 and 2 years	.02396
Kokinda (1989), 1 and 2 years	.11520
Kokinda (1989), 1 and 3 years	.11979
Kokinda (1989), 2 and 3 years	.11211
Melvin (1996), Fr–So and Fr–Jr	.01094
Melvin (1996), Fr–So and Fr–Sr	.01101
Melvin (1996), Fr–Jr and Fr–Sr	.01129
Mines, King, Hood, and Wood (1990), Cornell *Z* and Watson-Glaser	.07501
Shin, Lee, Ha, and Kim (2006), 1 and 2 years	.03129
Shin et al. (2006), 1 and 3 years	.03109
Shin et al. (2006), 2 and 3 years	.03123
Whitmire (2001), 1.5 and 3.5 years	.00096

Note. CCTST = California Critical Thinking Skills Test; CCTDI = California Critical Thinking Disposition Inventory; Fr = freshman; So = sophomore; Jr = junior; Sr = senior. σ_{d1d2} = estimated sampling covariance between the two effect size estimates.

points per point in time (e.g., 50 freshmen and 50 sophomores). By contrast, 100 participants in a longitudinal study provide 100 data points at each time. As such, reporting a single overall sample size that combines cross-sectional and longitudinal studies is not meaningful. To facilitate interpretation, Table 3 presents the sample sizes at both time points used to calculate each effect size. The sample of studies measuring critical thinking skills included approximately 16,185 longitudinal participants[2] and 9,392 cross-sectional participants. The sample of studies measuring critical thinking dispositions included 475 longitudinal participants

TABLE 5

Meta-analysis of critical thinking skill studies

Moderator	γ	95% confidence interval, lower	95% confidence interval, upper
Intercept	0.8834***	0.5852	1.1816
Time frame	0.0181	−0.0746	0.1108
Time frame^2	0.0249**	0.0061	0.0436
Longitudinal	−0.2703*	−0.5009	−0.0397
publication year	−0.0184***	−0.0248	−0.0121
σ^2_{study}	0.1165		

Note. Values are unstandardized coefficients from a mixed-effects meta-analysis. σ^2_{study} = variance component for the random effect of Study ID (effect sizes grouped within study).
**p* < .05. ***p* < .001. ****p* < .0001.

and 3,545 cross-sectional participants. There was some overlap between the skill and disposition samples, as three studies included scores on both constructs from the same group of participants.

The first author coded all studies in the meta-analysis. However, the second author independently coded a randomly selected subset of 15 studies (21% of the total) to ensure accuracy. There were no major discrepancies found, although one disagreement about a time frame estimate (1.5 years vs. 2 years) did occur. The resulting change did not substantially affect any of our conclusions.

Results

We used mixed-effects meta-analysis to model the effects of four moderator variables: time frame (0.5–4 years), study design (cross-sectional = 0, longitudinal = 1), sample[3] (nonnursing = 0, nursing = 1), and year of publication (1963–2011). For ease of interpretation and to allow the model to converge, we recentered the publication year variable by subtracting the earliest year in our sample from all values. Conceptually, this means that a value of zero for publication year corresponds to the year 1963. To test for nonlinear gains in critical thinking over time (e.g., larger or smaller gains during the entirety of college than would be expected from rescaling effect sizes from a single semester), we also included a quadratic term for time frame. Initial analyses indicated that the nursing/nonnursing moderator did not reach statistical significance. Additionally, a reduced model without this moderator produced lower values of the Aikake information criterion and Bayesian information criterion, indicating better fit. Therefore, we present results for the reduced model in the interest of simplicity.

Results are presented for changes in critical thinking skills across moderators in Table 5. We found evidence for significant moderator effects, $Q(4) = 9325.1340$, $p < .0001$, as well as significant residual heterogeneity, $Q(105) = 716.1532$, $p < .0001$. As expected, gains were larger across longer time frames. However, only the quadratic effect of time frame reached statistical significance. This finding suggests that gains in critical thinking skills during college are nonlinear, with the

TABLE 6

Meta-analysis of critical thinking skill studies including starting year in school

Moderator	Model 1	Model 2	Model 3
Intercept	1.1201***	1.0309	0.7893***
Time frame	−0.0069	−0.0101	0.1467***
Time frame^2	0.0313**	0.0320**	—
Starting year	—	0.073	0.0591
Longitudinal	−0.2167	−0.2358	−0.2077
Publication year	−0.0263***	−0.0267***	−0.0235***
σ^2_{study}	0.1622	0.1655	0.1496

Note. Values are unstandardized coefficients from three mixed-effects meta-analysis models. These analyses use a subset of 76 effect sizes for which a specific starting year in school could be identified. σ^2_{study} = variance component for the random effect of Study ID (effect sizes grouped within study).
**p < .001. ***p < .0001.

rate of change increasing over larger time intervals. An important caveat when interpreting this finding is that the time frames in many studies could not be attributed to specific years of college. For example, studies of gains during semester-long courses combine data from freshmen, sophomores, juniors, and seniors enrolled in those particular courses. As a result, the significant quadratic term is not a clear indicator that critical thinking gains accelerate in the later years of college.

To follow up on this possibility, we analyzed a subset of 76 effect sizes for which we could code a specific starting year (e.g., freshman = 1). In this subset, we compared three models: the original model, the original model with starting year added as an additional moderator, and the starting year added model with the quadratic time frame term removed. The purpose of removing the quadratic term was to determine if it was accounting for variance in effect sizes that would otherwise be attributed to starting year. Coefficients from all three models are presented in Table 6. The effect of starting year did not reach significance in either Model 2 or Model 3, suggesting that there may not be differential gains across different years of college.

We also analyzed differences between longitudinal and cross-sectional studies in our full sample. Our results suggest that study design has a substantial influence on effect size estimates. Controlling for other moderators, using a longitudinal design as opposed to a cross-sectional design was associated with a reduction of 0.27 *SD*s in estimated gains on critical thinking skills.

To better represent the joint effects of time frame and study design, we calculated model-predicted values for different levels of each moderator with publication year set to the sample mean (approximately 1994). The results in Table 7 show substantially larger effects for cross-sectional studies and longer time frames. We also predicted effect sizes for a hypothetical mixed sample with an equal number of cross-sectional and longitudinal studies. This analysis produced

TABLE 7

Model-predicted values for time frame and study design moderators with changes in critical thinking skills as the outcome

Time frame (years)	Cross-sectional *d*	Longitudinal *d*	Mixed *d*
0.5	N/A	0.00 [−0.44, 0.44]	N/A
1	0.30 [−0.16, 0.76]	0.03 [−0.41, 0.47]	0.16 [−0.28, 0.61]
2	0.39 [−0.07, 0.85]	0.12 [−0.32, 0.57]	0.26 [−0.19, 0.70]
3	0.54 [0.08, 0.99]	0.26 [−0.18, 0.71]	0.40 [−0.04, 0.84]
4	0.73 [0.27, 1.19]	0.46 [0.01, 0.90]	0.59 [0.15, 1.04]

Note. N/A = not applicable. Values in brackets are 80% credibility intervals. Mixed *d* is calculated using a value of 0.5 for the study design moderator (cross-sectional = 0, longitudinal = 1). N/A values are used to avoid extrapolating beyond the meta-analytic sample as there were no cross-sectional single-semester studies. The publication year moderator was fixed at its sample mean for these analyses.

a 4-year gains estimate of 0.59 *SD*s, as opposed to 0.73 for cross-sectional studies and 0.46 for longitudinal designs.

The compilation of studies spanning 48 years also allowed us to examine changes in critical thinking gains over time. To test the suggestion that college has become less effective at teaching critical thinking, we included publication year as a moderator. Our moderator analysis provides some support for this notion. Holding other moderators constant, more recent studies provided significantly smaller effect sizes than older studies. Given an equal mix of cross-sectional and longitudinal studies, the predicted 4-year gain is 1.22 *SD*s for a study published in 1963 (80% credibility interval [0.75, 1.68]), whereas the predicted gain is only 0.33 for a study published in 2011 (80% credibility interval [−0.11, 0.78]).

Our search revealed a relatively small sample of studies measuring the disposition toward thinking critically, all of which used the CCTDI. Given the small meta-analytic sample, we tested simpler moderator analyses excluding sample type and publication year. Using Aikake information criterion and Bayesian information criterion as criteria, a simple model with only the linear effect of time frame produced the best fit. This model is presented in Table 8. As expected, longer time frames were associated with larger effect sizes, $Q(1) = 22.8770$, $p < .0001$. However, significant residual heterogeneity remained after accounting for this effect, $Q(12) = 52.1465$, $p < .0001$. Predicted values for different time frames are shown in Table 9. We estimate an average gain of 0.55 *SD*s on the disposition toward critical thinking over 4 years of college.

As previously mentioned, we imputed *SD*s from other samples for several studies in which no *SD*s were presented. It is possible that our specific decisions about the sources of imputed values could have affected the overall results of this study. To address this concern, we performed a sensitivity analysis using alternate imputed values. Specifically, we computed new imputed values using *N*-weighted averages of pretest *SD*s from the meta-analytic sample. This resulted in three new

TABLE 8

Meta-analysis of critical thinking disposition studies

Moderator	γ	95% confidence interval, lower	95% confidence interval, upper
Intercept	0.1444	−0.0534	0.3422
Time frame	0.1015***	0.0599	0.1431
σ^2_{study}	0.0555		

Note. Values are unstandardized coefficients from a mixed-effects meta-analysis. σ^2_{study} = variance component for the random effect of Study ID (effect sizes grouped within study).
***$p < .0001$.

TABLE 9

Model-predicted values for different time frames with changes in critical thinking dispositions as the outcome

Time frame (years)	*d*	80% credibility value, lower	80% credibility value, upper
0.5	0.20	−.13	.52
1	0.25	−.08	.57
2	0.35	.02	.67
3	0.45	.12	.77
4	0.55	.22	.88

imputed values for the Watson-Glaser, California Critical Thinking Skills Test (CCTST), and CCTDI. For the studies in Table 2, we replaced our original imputed values with these new values. We then reran the mixed-effects meta-analyses for critical thinking skills and dispositions using the new effect sizes and sampling variances. The new values did not affect our model selection decisions and had trivial effects on the magnitudes of moderator coefficients (see Tables A1 and A2 in Appendix A for these results). Thus, our overall conclusions appear to be robust to changes in imputed *SD*s.

Discussion

Our study suggests that students make substantial gains in critical thinking during college. We estimate the overall effect of college on critical thinking skill at 0.59 *SD*s, which paints a slightly more optimistic picture than recent estimates by Arum and Roksa (2011) and Pascarella and Terenzini (2005). However, our overall findings are fairly consistent with these studies. Arum and Roksa estimate a gain of 0.18 *SD*s over three semesters, which falls well within our 80% credibility intervals for both 1-year and 2-year effect sizes. Their estimate of 0.47 *SD*s over 4 years is nearly identical to our point estimate of 0.46 for 4-year longitudinal studies. Similarly, Pascarella and Terenzini's overall estimate of 0.50 *SD*s from a combination of longitudinal and cross-sectional studies is reasonably similar to our mixed designs estimate of 0.59 (especially considering that our effect size estimates are corrected for unreliability whereas theirs are not). It is worth noting that a 0.50 *SD* gain for person who starts at the 50th percentile would lift him or her to the 69th percentile, no small improvement in our minds.

A major contribution of the present study is that we analyze time frame as a moderator rather than aggregating all time frames into a single estimate, providing more information about nonlinear patterns in the growth rate. Our quadratic analysis of the time frame moderator suggests that the rate of gains in critical thinking skills increases across larger time frames. This finding suggests that it may be inappropriate to rescale and collapse effect sizes from different time frames (e.g., Ortiz, 2007). However, follow-up analyses failed to link this effect to specific years in college. Thus, we did not find unequivocal support for an *acceleration effect* where critical thinking gains become more rapid in the later years of college. On the other hand, our results also do not support Arum and Roksa's (2011) suggestion that critical thinking may increase more in the early stages of college.

Another benefit of our study is that it includes cross-sectional designs, which were excluded or grouped with longitudinal studies in other reviews. Considering these designs is important because they make up a large part of the literature. However, treating them as equivalent to longitudinal designs may lead to erroneous conclusions; we found that cross-sectional studies produced substantially larger effect sizes than longitudinal studies. This finding suggests that critical thinking researchers should carefully consider the effects of their study design on the final results. Cross-sectional studies may be confounded if students who score higher on critical thinking tests are more likely to remain in college. The result would be inflated effect size estimates, as low-performing students contribute only to the pretest mean before dropping out. Since critical thinking is related to college performance (Kuncel, 2011), this confound is a possibility.

On the other hand, longitudinal studies suffer from a similar self-selection problem. Longitudinal samples are typically restricted to students who remain in college for both data collections. As a result, effect size estimates may be downwardly biased by range restriction, a statistical artifact that results from artificially reduced variance in the outcome of interest (e.g., Bobko et al., 2001). In addition, students who initially score low on critical thinking may simply have more room to improve than their high-ability counterparts. The degree to which this is problematic largely depends on the research question of interest. If a researcher wants to know the effect of college on students who stay in college, then the range restriction issue is not a concern. However, it should be recognized that longitudinal effect sizes are likely to be larger if the dropout rate is reduced.

Another somewhat worrisome finding is that observed gains in critical thinking appear to have deteriorated over time despite increased interest in fostering critical thinking skills. This finding is not conclusive evidence that college educations have declined in quality because there are a number of possible explanations for the observed effect. First, it is worth noting that a large amount of residual variance in effect sizes remained after our moderators were accounted for. Thus, publication year may be acting as a proxy for some missing variable that also changed over time. Such a variable could either directly affect gains on critical thinking or affect the measurement of gains. For example, changes in curricula or student behaviors could bring about reductions in the effectiveness of college (e.g., Terenzini et al., 1995). On the other hand, changes in study design or the quality of studies might affect observed effect sizes in ways not accounted for by our single design moderator. Such differences would be artifactual rather than representing true changes in how well critical thinking is learned. It is worth noting that our estimated 4-year gain for a study published in 1963 is 1.22 *SD*s. Such an effect size is perhaps suspiciously large and may partially reflect lower standards for research design and implementation.

Another potential explanation is that students are now coming to college with a reduced readiness to learn critical thinking skills. One reason for this may be that students have increasingly learned more critical thinking skills before entering college. If the skills taught in college are already present in a greater proportion of students, then overall gain scores should be reduced. Alternatively, college attendance has increased over time, and many new students may not be sufficiently prepared to learn more complex reasoning skills. A final possibility is that students have become less willing or able to learn critical thinking skills over time. As of now, these possibilities are mere speculation. Further research is needed to determine the true causes of this phenomenon.

Implications

Although college education may lag in other ways, it is not clear that more time and resources should be invested in teaching domain-general critical thinking. For a specific group of individuals who already possess above-average cognitive abilities, a gain of 0.59 *SD*s on a purportedly general ability is quite impressive (comparable to going from the 50th percentile to the 72nd percentile). The effect of college on critical thinking is larger than the average effect of educational variables on academic achievement (0.40 *SD*) and even rivals the effect of disposition toward learning (0.61 *SD*; Hattie, 1992). Put differently, college appears to produce critical thinkers about as well as motivation produces good students.

College also appears to foster more favorable dispositions toward critical thinking. As an attitudinal construct, critical thinking disposition is arguably even less trainable than critical thinking skill. However, it does appear that the college experience can have a substantial impact. Our 4-year gain estimate of 0.55 *SD*s is not markedly different from Pascarella and Terenzini's (2005) estimate of 0.50 (although there is some overlap between our disposition samples). This finding is particularly important because critical thinking disposition may be the only domain-general form of critical thinking, in that a willingness or desire to question and critique is clearly applicable across settings. The average increase of over half an *SD* on this general disposition is encouraging.

When considering educational interventions, the amount of value added relative to other potential investments must be a central consideration. Time spent teaching critical thinking is time not spent teaching other things, such as reading, writing, mathematics, and profession-specific knowledge. If our efforts to foster critical thinking are inadequate, then the same surely holds for these domains in which observed gains are similar to gains in critical thinking. Pascarella and Terenzini (2005) estimate 4-year gains of 0.77 *SD*s, 0.62 *SD*s, and 0.55 *SD*s for English, science, and mathematics skills, respectively, which are similar to our estimates for critical thinking gains. It is also clear that there is significant room for improvement in these fundamental competencies. Students in the United States score only around the average among Organization for Economic Co-operation and Development (2010) member countries on reading and science skills, and they are actually below average at mathematics.

Given the skill deficiencies that exist in multiple areas within the labor force (Galagan, 2010), we must carefully consider where we invest educational resources. It is unlikely that additional investment in domain-general critical thinking will provide a solution to our problems. Our analysis of nursing samples failed to find any long-term advantage of the NLN's critical thinking requirement; nursing students simply did not improve more than their nonnursing counterparts. Although Abrami et al. (2008) found an average effect size of 0.34 for critical thinking interventions, the nursing data suggest that such interventions may ultimately have little incremental impact above and beyond the gains that naturally occur over the span of college.

Limitations and Future Directions

The central limitation of the literature we synthesize is the inability to make clear causal conclusions, a limitation that is problematic in two ways. First, the studies reviewed do not distinguish the effects of college from ordinary maturation effects, a persistent problem in this body of research (Pascarella & Terenzini, 1991). There is some evidence that the observed effects are largely due to college. Pascarella and Terenzini (1991, 2005) reviewed studies that control for maturation and other confounds, and they concluded that college still produces significant effects. However, it may still be the case that critical thinking increases naturally with age and that some of the observed changes occur independently of college education. Of course, a true experimental design is still lacking in the literature. Since it is not realistic to randomly assign participants to attend or not attend college, a degree of uncertainty is likely to persist.

A second issue is that nursing differs from other majors in ways other than attention to critical thinking. One must consider the possibility that other aspects of nursing education or nursing students could cause differential gains in critical thinking skills. If some unknown feature of nursing programs were to suppress critical thinking, then the effect of the curricular difference might be masked. In other words, nursing programs' additional focus on critical thinking could actually be alleviating a relative deficit that would otherwise exist. Therefore, we cannot rule out the possibility that a critical thinking curriculum *could* produce lasting incremental gains. At the very least, our study simply shows that such gains—if they exist—are not readily apparent despite a large literature devoted to searching for them. As previously mentioned, it is not highly feasible to randomly assign participants to a long-term critical thinking curriculum. Therefore, the analysis of nonequivalent comparison groups is arguably the best evidence that is presently available.

The evidence from nursing samples casts doubt on the amount of value added by explicitly training domain-general critical thinking in college. The critical thinking literature could benefit from a change in focus to incorporate domain-specific critical thinking. Our literature search revealed relatively few studies that measured critical thinking in a specific content domain and even fewer that compared specific and general measures. Most of those we did find were tests of critical thinking in nursing or psychology. Without sufficient studies of critical thinking in other domains, we are unable to draw generalizable conclusions about changes in critical thinking skills. If the logic items on the Watson-Glaser and other common tests represent only one domain out of many in which one can think critically, then the current literature has largely ignored a crucial aspect of critical thinking.

It is plausible that domain-specific measures would show stronger gains in college and track better with important outcomes. For example, Renaud and Murray (2008) compared gains on domain-general and domain-specific critical thinking tests following a brief intervention. Students in both experimental and control conditions read a passage about personality theory. The experimental group then completed critical thinking questions about the passage, whereas the control group completed simple recall questions. Participants in the experimental group showed larger gains in domain-specific, but not domain-general, critical thinking. Another experiment by Williams, Oliver, and Stockdale (2004) indicated that psychology students showed significant gains on a measure of psychological critical thinking when critical thinking practice was incorporated into the course. However, there were not significant gains on the Watson-Glaser for either experimental or control groups. Williams (2003) found that students in an educational psychology course showed more significant gains on psychological critical thinking than domain-general critical thinking. In particular, students who received high grades in the course improved more on the domain-specific test. This finding suggests that changes in domain-specific critical thinking may be related to mastery of that domain.

The domain-specificity hypothesis could also explain the failure of nursing programs to produce larger gains in critical thinking through explicit instruction. The type of critical thinking infused into current nursing curricula may not be captured well by traditional measures. General critical thinking inventories measure the ability to employ a specific set of logical rules, but these rules are not necessarily the ones used to think critically about the condition of a patient or which treatment is appropriate. Nursing curricula may focus on teaching critical thinking rules that are useful to nurses but not as useful for increasing scores on the Watson-Glaser. Alternatively, nursing education may not be conducive to retention of the skills taught in a general critical thinking lecture. Skills are typically more likely to be retained if they are practiced (Campbell & Kuncel, 2001), and it is unlikely that the day-to-day experience of nursing education affords much explicit practice at recognizing post hoc fallacies or using *modus ponens*. The domain-specificity hypothesis suggests that critical thinking skills taught in one domain (e.g., formal logic) are unlikely to transfer well to another (e.g., nursing; McPeck, 1984). Under this paradigm, it is unsurprising that nursing students would fail to apply (and thus retain) the ability to analyze formal arguments.

The retention and application of critical thinking gains beyond college also require further study. The present study demonstrates that college students learn critical thinking skills, but this does not guarantee that they retain these skills long after college or apply them in other contexts (Campbell & Kuncel, 2001; Schmidt & Bjork, 1992). Our search did not reveal any studies that followed up with college graduates to determine their levels of critical thinking skill or disposition later in life. If critical thinking skills are not practiced as frequently after graduation, they may diminish over time. Similarly, it is likely that the disposition toward critical thinking is influenced by the norms and environment of college, which generally promote open-mindedness and other intellectual virtues. It may be the case that a departure from the college environment would be associated with regression toward the mean of critical thinking dispositions. To the best of our knowledge, the above possibilities have never been formally tested.

Conclusion

Our initial results argue against investing additional time and resources in teaching domain-general critical thinking. Although the set of specific skills measured by critical thinking tests is important, spending more time on them involves trade-offs with other important skills. The evidence suggests that basic competencies such as reading and mathematics are more amenable to improvement beyond the gains currently observed, and the need is arguably more desperate. In addition, critical thinking in major-related domains may be a more practical target for instruction than the kind of critical thinking measured by domain-general tests, although further research is needed to explore this possibility. Regardless, our findings should not be a cause for pessimism about the future of critical thinking in higher education. On the contrary, the present study has demonstrated that college is already effective at fostering critical thought, leaving more resources free to pursue other educational goals.

Appendix A

Sensitivity Analyses

TABLE A1

Meta-analysis of critical thinking skill studies using alternate imputed standard deviations

Moderator	γ	95% confidence interval, lower	95% confidence interval, upper
Intercept	0.8818***	0.5846	1.1790
Time frame	0.0184	−0.0743	0.1111
Time frame^2	0.0248**	0.0060	0.0435
Longitudinal	−0.2725*	−0.5020	−0.0429
Publication year	−0.0184***	−0.0247	−0.0121
σ^2_{study}	0.1151		

TABLE A2

Meta-analysis of critical thinking disposition studies using alternate imputed standard deviations

Moderator	γ	95% confidence interval lower	95% confidence interval upper
Intercept	0.1393	−0.0533	0.3318
Time frame	0.1013***	0.0598	0.1428
σ^2_{study}	0.0509		

Notes

[1]It is worth noting that although the authors provide "best estimates" for these effect sizes, their exact methodology for arriving at these estimates is unclear. They appear to use meta-analytic techniques (Pascarella & Terenzini, 2005, pp. 12, 150). However, they provide a narrative review of several cross-sectional and longitudinal studies and then simply report an overall effect size estimate. Some of these studies do not contain the information necessary to compute an effect size, but the authors attempt to estimate an effect size regardless (p. 157).

[2]Two longitudinal studies had unequal sample sizes at different times due to incomplete data for some participants (Ewen, 2001; Johnson, 2002). For these studies, we used an average of the sample sizes at Time 1 and Time 2 to compute the overall sample size estimate.

[3]The nonnursing category includes the following majors and categories of major: unspecified/mixed, humanities, liberal arts, science, mathematical and social sciences, social sciences, health science, business, psychology, engineering, architecture/architecture engineering, physical therapy, and prehealth science.

References

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

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, *78*, 1102–1134. doi:10.3102/0034654308326084

*Adams, M. H., Stover, L. M., & Whitlow, J. F. (1999). A longitudinal evaluation of baccalaureate nursing students' critical thinking abilities. *Journal of Nursing Education*, *38*, 139–141.

Adams, M. H., Whitlow, J. F., Stover, L. M., & Johnson, K. W. (1996). Critical thinking as an educational outcome: An evaluation of current tools of measurement. *Nurse Educator*, *21*(3), 23–32. doi:10.1097/00006223-199605000-00009

*Arburn, T., & Bethel, L. (1999, March). *Assisting at risk community college students' acquisition of CT learning strategies*. Paper presented at the National Association for Research in Science Teaching conference, Boston, MA.

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

*Bartlett, D. J., & Cox, P. D. (2002). Measuring change in students' critical thinking ability: Implications for health care education. *Journal of Allied Health*, *31*, 64–69.

*Beckie, T. M., Lowry, L. W., & Barnett, S. (2001). Assessing critical thinking in baccalaureate nursing students: A longitudinal study. *Holistic Nursing Practice*, *15*(3), 18–26. doi:10.1097/00004650-200104000-00006

*Berger, M. C. (1984). Critical thinking ability and nursing students. *Journal of Nursing Education*, *23*, 306–308.

*Blaich, C., & Wise, K. (2008). *Overview of findings from the first year of the Wabash National Study of Liberal Arts Education*. Crawfordsville, IN: Center of Inquiry in the Liberal Arts at Wabash College.

*Blaich, C., & Wise, K. (2011). *How do students change over four years of college?* Crawfordsville, IN: Center of Inquiry in the Liberal Arts at Wabash College.

Bobko, P., Roth, P. L., & Bobko, C. (2001). Correcting the effect size of *d* for range restriction and unreliability. *Organizational Research Methods*, *4*, 46–61. doi:10.1177/109442810141003

*Brigham, C. F. (1989). *Critical thinking skills in nursing students progressing through a nursing curriculum* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8918496)

*Burgess, J. A. M. (2003). *The impact of the school of nursing curriculum on generic baccalaureate nursing students' ability to think critically at Austin Peay State University* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3094628)

Campbell, J. P., & Kuncel, N. R. (2001). Individual and team training. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work & organizational psychology: Vol. 1. Personnel psychology* (pp. 278–312). Thousand Oaks, CA: Sage.

*Colbert, K. R. (1987). *The effects of CEDA and NDT debate training on critical thinking ability* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8626790)

*Colucciello, M. L. (1997). Critical thinking skills and dispositions of baccalaureate nursing students—a conceptual model for evaluation. *Journal of Professional Nursing*, *13*, 236–245. doi:10.1016/S8755-7223(97)80094-4

*Criner, L. A. (1992). *Teaching thinking and reasoning: A study of critical thinking in adults* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9237328)

*Dale, P. M., Ballotti, D., Handa, S., & Zych, T. (1997). An approach to teaching problem solving in the classroom. *College Student Journal*, *31*, 76–79.

*Daly, W. M. (2001). The development of an alternative method in the assessment of critical thinking as an outcome of nursing education. *Journal of Advanced Nursing*, *36*, 120–130. doi:10.1046/j.1365-2648.2001.01949.x

DeAngelo, L., Hurtado, S., Pryor, J. H., Kelly, K. R., Santos, J. L., & Korn, W. S. (2009). *The American college teacher: National norms for the 2007-2008 HERI faculty survey*. Los Angeles, CA: Higher Education Research Institute.

Drouin, L. E. (1992). An investigation of the critical thinking ability of engineering students seeking a bachelor of science degree. *Dissertation Abstracts International*, *53*, 1070A.

*Erickson, R. L. (1999). *The effects of database training and use on the development of critical-thinking skills* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9955755)

*Ewen, L. L. (2001). *A longitudinal study of nursing students' self-esteem, locus of control, anxiety, critical thinking and academic achievement* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3041104)

*Facione, N. C., & Facione, P. A. (1997). *Critical thinking assessment in nursing education programs: An aggregate data analysis*. Millbrae: California Academic Press.

Facione, N. C., Facione, P. A., & Sanchez, C. A. (1994). Critical thinking disposition as a measure of competent clinical judgment: The development of the California Critical Thinking Disposition Inventory. *Journal of Nursing Education*, *33*, 345–350.

Facione, P. A. (2011). *Critical thinking: What it is and why it counts*. Millbrae: California Academic Press.

*Fleeger, R. L. (1986). *Critical thinking and moral reasoning behavior of baccalaureate nursing students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8616530)

*Forbes, C. A. (1997). *Analyzing the growth of the critical thinking skills of college calculus students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9737709)

*Frost, S. H. (1991). Fostering the critical thinking of college women through academic advising and faculty contact. *Journal of College Student Development*, *32*, 359–366.

Galagan, P. (2010). *Bridging the skills gap*. Alexandria, VA: American Society for Training and Development.

*Giancarlo, C. A., & Facione, P. A. (2001). A look across four years at the disposition toward critical thinking among undergraduate students. *Journal of General Education*, *50*, 29–55. doi:10.1353/jge.2001.0004

Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 357–376). New York, NY: Russell Sage Foundation.

*Goodin, H. I. J. (2005). *The use of deliberative discussion as a teaching strategy to enhance the critical thinking abilities of freshman nursing students* (Doctoral dissertation). Retrieved from https://etd.ohiolink.edu/ap/10?214682536937066::NO:10:P10_ETD_SUBID:63853

*Gross, Y., Takazawa, E., & Rose, C. (1987). Critical thinking and nursing education. *Journal of Nursing Education*, *26*, 317–323.

*Hagedorn, L. S., Pascarella, E. T., Edison, M., Braxton, J., Nora, A., & Terenzini, P. T. (1999). Institutional context and the development of critical thinking: A research note. *Review of Higher Education*, *22*, 265–286. doi:10.1353/rhe.1999.0010

Halpern, D. F. (2001). Assessing the effectiveness of critical thinking instruction. *Journal of General Education*, *50*, 270–286. doi:10.1353/jge.2001.0024

*Hancock, B. W. (1981). *The effect of guided design on the critical thinking ability of college level administrative science students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8206458)

Hattie, J. A. (1987). Identifying the salient facets of a model of student learning: A synthesis of meta-analyses. *International Journal of Educational Research*, *11*, 187–212.

Hattie, J. A. (1992). Measuring the effects of schooling. *Australian Journal of Education*, *36*, 5–13.

Hill, K. E. (1995). Critical thinking and its relation to academic, personal, and moral development in the college years. *Dissertation Abstracts International*, *56*, 4603B.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

*Isaacs, L. G. (1990). *A study of the effect of teaching critical thinking in an introductory nursing course* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9104407)

Jacobs, S. S. (1995). Technical characteristics and some correlates of the California Critical Thinking Skills Test, Forms A and B. *Research in Higher Education*, *36*, 89–108. doi:10.1007/BF02207768

*Johnson, F. E. (2002). *Critical thinking skills of baccalaureate nursing students in a collaborative nursing program*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3041691)

*Jones, J. H., & Morris, L. V. (2007). Evaluation of critical thinking skills in an associate degree nursing program. *Teaching and Learning in Nursing*, *2*, 109–115. doi:10.1016/j.teln.2007.07.006

*Keller, R. (1993). *Effects of an instructional program on critical thinking and clinical decision-making skills of associate degree nursing students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9404905)

*Kintgen-Andrews, J. (1988). *Development of critical thinking: Career ladder PN and AD nursing students, pre-health science freshmen, generic baccalaureate sophomore nursing students*. Minneapolis: The University of Minnesota. Retrieved from ERIC database. (ED297153)

Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). The collegiate learning assessment: Facts and fantasies. *Evaluation Review*, *31*, 415–439. doi:10.1177/0193841X07303318

Koenig, J., Herman, J., Duncan, G., Knapp, D., Kyllnonen, P., Sackett, P., . . . Iverson, K. (2011). *Assessing 21st century skills: Summary of a workshop*. *Washington, DC*: National Academies Press.

*Kokinda, M. A. (1989). *The measurement of critical thinking skills in a selected baccalaureate nursing program* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9004798)

Kuncel, N. R. (2011). *Measurement and meaning of critical thinking* (Research report for the NRC 21st Century Skills Workshop). Washington, DC: National Research Council.

*Lehmann, I. J. (1963). Changes in critical thinking, attitudes, and values from freshman to senior years. *Journal of Educational Psychology*, *54*, 305–315. doi:10.1037/h0045302

*Leppa, C. J. (1997). Standardized measures of critical thinking: Experience with the California critical thinking tests. *Nurse Educator*, *22*, 29–33. doi:10.1097/00006223-199709000-00012

*Lewis, D. R., & Dahl, T. (1972). Critical thinking skills in the principles course: An experiment. In A. Welsh (Ed.), *Research papers in economic education* (pp. 50–68). New York, NY: Joint Council on Economic Education.

*Lyons, E. M. (2008). Examining the effects of problem-based learning and NCLEX-RN scores on the critical thinking skills of associate degree nursing students in a southeastern community college. *International Journal of Nursing Education Scholarship*, *5*, 1–17. doi:10.2202/1548-923X.1524

*McCarthy, P., Schuster, P., Zehr, P., & McDougal, D. (1999). Evaluation of critical thinking in a baccalaureate nursing program. *Journal of Nursing Education*, *38*, 142–144.

*Mcdonough, M. F. (1997). *An assessment of critical thinking at the community college level* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9738997)

*McKee, M. L. (2006). *Overcoming obstacles in teaching college students how to think critically: The effects of an instructional intervention* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3222188)

McMillan, J. H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education*, *26*, 3–29. doi:10.1007/BF00991931

McPeck, J. E. (1984). Stalking beasts, but swatting flies: The teaching of critical thinking. *Canadian Journal of Education*, *9*, 28–44. doi:10.2307/1494448

McPeck, J. E. (1990). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher*, *19*(4), 10–12. doi:10.3102/0013189X019004010

*Meiss, G. T., & Bates, G. W. (1984, May). *Cognitive and attitudinal effects of reasoning message strategies*. Paper presented at the annual conference of the International Communication Association, San Francisco, CA.

*Melvin, T. M. (1996). *Evaluation of curriculum effectiveness in developing critical thinking skills at Wentworth Institute of Technology* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9720530)

Mentkowski, M., & Rogers, G. (1985, October). *Longitudinal assessment of critical thinking in college: What measures assess curricular impact?* Paper presented at the annual meeting of The Mid-Western Educational Research Association, Chicago, IL.

*Miller, M. H. (1987). *Impact of a baccalaureate registered nurse program on the critical thinking skills of students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8808292)

*Mines, R. A., King, P. M., Hood, A. B., & Wood, P. K. (1990). Stages of intellectual development and associated critical thinking skills in college students. *Journal of College Student Development*, *31*, 538–547.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105–125. doi:10.1037//1082-989X.7.1.105

*Nathan, Y. H. (1997). *Critical thinking: Impact on two classes of nursing students in an academic year* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9734083)

*Norwood, B. R. (1998). *Critical thinking in nursing students: A comparison of designs*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9903935)

*Notarianni, M. A. (1991). *An investigation of the critical thinking ability of associate and baccalaureate degree nursing students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9129260)

Organisation for Economic Co-operation and Development. (2010). *PISA 2009 results: Executive summary*. Retrieved from www.oecd.org/pisa/pisaproducts/46619703.pdf

Ortiz, C. M. A. (2007). *Does philosophy improve critical thinking skills?* (Unpublished master's thesis). The University of Melbourne, Victoria, Australia.

*Pascarella, E. T. (1989). The development of critical thinking: Does college make a difference? *Journal of College Student Development*, *30*, 19–26.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco, CA: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research*. San Francisco, CA: Jossey-Bass.

Peterson, N. G., Mumford, M., Borman, W., Jeanneret, P., Fleishman, E., & Levin, K. (1997). *O*NET final technical report*. Salt Lake City: Utah Department of Workforce Services.

*Pierce, T. (1994). *Institutional effectiveness assessment process, 1993-94. Executive summary*. Seattle, WA: South Seattle Community College. Retrieved from ERIC database. (ED381223)

*Pierce, T. (2011). *Introductory computer programming courses used as a catalyst to critical thinking development* (Doctoral dissertation). Retrieved from http://cardinalscholar.bsu.edu/handle/123456789/195140

Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., . . . Downer, J. (2012). *Third grade follow-up to the Head Start impact study*. Washington, DC: U.S. Department of Health and Human Services. Retrieved from www.acf.hhs.gov/sites/default/files/opre/head_start_report.pdf

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Renaud, R. D., & Murray, H. G. (2008). A comparison of a subject-specific and a general measure of critical thinking. *Thinking Skills and Creativity*, *3*, 85–93. doi:10.1016/j.tsc.2008.03.005

*Richards, M. A. (1977). One integrated curriculum: An empirical evaluation. *Nursing Research*, *26*, 90–95. doi:10.1097/00006199-197703000-00003

*Rose, M. M. (1997). *Critical thinking skills instruction for postsecondary students with and without learning disabilities: The effectiveness of icons as part of a literature curriculum* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9806188)

*Ross, G., & Semb, G. (1981). Philosophy can teach critical thinking skills. *Teaching Philosophy*, *4*, 111–122. doi:10.5840/teachphil19814233

*Ruff, L. G. (2005). *The development of critical thinking skills and dispositions in first-year college students: Infusing critical thinking instruction into a first-year transitions course* (Doctoral dissertation). University of Maryland, College Park. Retrieved from http://drum.lib.umd.edu/handle/1903/3155

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *3*, 207–217. doi:10.1111/j.1467-9280.1992.tb00029.x

*Seifert, T. A., Pascarella, E. T., Colangelo, N., & Assouline, S. G. (2007). The effects of honors program participation on experiences of good practices and learning outcomes. *Journal of College Student Development*, *48*, 57–74. doi:10.1353/csd.2007.0007

*Shin, K. R., Lee, J. H., Ha, J. Y., & Kim, K. H. (2006). Critical thinking dispositions in baccalaureate nursing students. *Journal of Advanced Nursing*, *56*, 182–189. doi:10.1111/j.1365-2648.2006.03995.x

*Smith, M. L. (1995). *A quantitative analysis of critical thinking abilities, learning and study strategies, and academic achievement in associate degree nursing students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9602082)

*Solon, T. (2001). Improving critical thinking in an introductory psychology course. *Michigan Community College Journal: Research & Practice*, *7*, 73–80.

*Solon, T. (2003). Teaching critical thinking: The more, the better. *The Community College Enterprise*, *9*, 25–38.

*Solon, T. (2007). Generic critical thinking infusion and course content learning in introductory psychology. *Journal of Instructional Psychology*, *34*, 95–109.

*Spaulding, S. C., & Kleiner, K. A. (1992). The relationship of college and critical thinking: Are critical thinkers attracted or created by college disciplines? *College Student Journal*, *26*, 162–166.

*Sukspringarm, P. (1976). *An experimental study comparing the effects of BSCS and traditional biology on achievement, understanding of science, critical thinking ability, and attitude toward science of the first year students at the Sakon Nakorn Teachers' College, Thailand* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 7624805)

*Sullivan, E. J. (1987). Critical thinking, creativity, clinical performance, and achievement in RN students. *Nurse Educator*, *12*, 12–16. doi:10.1097/00006223-198703000-00008

Terenzini, P. T., Springer, L., Pascarella, E. T., & Nora, A. (1995). Influences affecting the development of students' critical thinking skills. *Research in Higher Education*, *36*, 23–39. doi:10.1007/BF02207765

*Thompson, C., & Rebeschi, L. M. (1999). Critical thinking skills of baccalaureate nursing students at program entry and exit. *Nursing and Health Care Perspectives*, *20*, 248–252.

*Tomlinson-Keasey, C., & Eisert, D. (1978). *Second year evaluation of the ADAPT program* (Essays from and about the ADAPT Program). Retrieved from http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1030&context=adaptessays

Tsui, L. (1998). *Fostering critical thinking in college students: A mixed-methods study of influences inside and outside of the classroom* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9917229)

*Turner, P. (2000). Critical thinking and intellectual development: Changes associated with baccalaureate nursing education. *Korean Journal of Thinking & Problem Solving*, *10*, 41–48.

Viechtbauer, W. (2014). *Metafor: Meta-analysis package for R* (Version 1.9-5) [Computer software]. Retrieved from https://cran.r-project.org/web/packages/metafor/index.html

Wang, Y., & Liao, H. C. (2012). The promotion of critical thinking in baccalaureate nursing English programs. *African Journal of Business Management*, *6*, 3188–3196.

*Watson, G., & Glaser, E. M. (1980). *Watson-Glaser critical thinking appraisal*. San Antonio, TX: Psychological Corporation.

*Wheeler, L. A., & Collins, S. K. R. (2003). The influence of concept mapping on critical thinking in baccalaureate nursing students. *Journal of Professional Nursing*, *19*, 339–346. doi:10.1016/S8755-7223(03)00134-0

*Whitmire, E. (2001). The relationship between undergraduates' background characteristics and college experiences and their academic library use. *College & Research Libraries*, *62*, 528–540.

*Williams, R. L. (2003). *Critical thinking as a predictor and outcome measure in a large undergraduate educational psychology course*. Knoxville: The University of Tennessee. Retrieved from ERIC database. (ED 478075)

*Williams, R. L., Oliver, R., & Stockdale, S. L. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *Journal of General Education*, *53*, 37–58. doi:10.1353/jge.2004.0022

*Winter, D. G., & McClelland, D. C. (1978). Thematic analysis: An empirically derived measure of the effects of liberal arts education. *Journal of Educational Psychology*, *70*, 8–16. doi:10.1037//0022-0663.70.1.8

*Yuan, H., Kunaviktikul, W., Klunklin, A., & Williams, B. A. (2008). Improvement of nursing students' critical thinking skills through problem-based learning in the people's republic of china: A quasi-experimental study. *Nursing & Health Sciences*, *10*, 70–76. doi:10.1111/j.1442-2018.2007.00373.x

Authors

CHRISTOPHER R. HUBER is a PhD student in the industrial/organizational psychology program at the University of Minnesota, 75 East River Parkway, Minneapolis, MN 55455; e-mail: *huber195@umn.edu*. His research has focused on the measurement, predictive power, and change of individual differences, with an emphasis on personality and standardized testing in education. He also studies psychometric and statistical issues such as dimension reduction techniques, modeling hierarchical data, and the use of forced choice scales to reduce faking.

NATHAN R. KUNCEL is the Marvin D. Dunnette Distinguished Professor of Psychology and a McKnight Presidential Fellow at the University of Minnesota, N218 Elliott Hall, 75 E. River Parkway, Minneapolis, MN 55455; e-mail: *kunce001@umn.edu*. He is also an associate professor in the Personality, Individual Differences and Behavioral Genetics area. He has received the Cattell Early Career Research Award from the Society of Multivariate Experimental Psychology and the Anne Anastasi Early Career Award from American Psychological Association's Division 5 (Evaluation, Measurement, & Statistics). His research is focused on the measurement of human cognitive and noncognitive characteristics and their relationship to important life outcomes, including academic and job performance. He has studied the power of individual differences for predicting different aspects of performance with a focus on the use of standardized tests (e.g., SAT, the GRE® test, GMAT®, PCAT, MAT), prior accomplishments (e.g., high school rank, extracurricular accomplishments), and personality traits for predicting academic performance in college and graduate school. He also conducts research on psychological measurement, response distortion, and meta-analytic methods. His research has appeared in *Science*, *Psychological Bulletin*, *Perspectives on Psychological Science*, *Review of Educational Research*, *Journal of Personality* and *Social Psychology*, and *Journal of Applied Psychology*, among others. He earned his doctorate from the University of Minnesota.