

Teaching Postsecondary Students to Use Analogies as a Cognitive Learning Strategy: An Intervention

Joseph C. Tise,^{1*} Rayne A. Sperling,² Michael S. Dann,³ and Taylor M. Young¹

¹Department of Teaching, Learning, and Culture, Texas A&M University, College Station, TX 77843;

²Department of Educational Psychology, Counseling, and Special Education and ³Department of Biology, The Pennsylvania State University, University Park, PA 16802

ABSTRACT

Analogical reasoning is an important type of cognition often used by experts across domains. Little research, however, has investigated how generating analogies can support college students' self-regulated learning (SRL) of biology. This study therefore evaluated a contextualized cognitive learning strategy intervention designed to teach students to generate analogies as a learning strategy to aid learning within a university biology course. Participants ($n = 179$) were taught how to generate analogies as a learning strategy to learn about plant and animal physiology. We hypothesized the quality of students' generated analogies would increase over time, and their analogical reasoning, knowledge of cognition (KOC; a component of metacognitive awareness), and course performance would be higher after intervention, controlling for associated pre-intervention values. Regression analyses and repeated-measures analysis of variance indicated a positive relationship between generated-analogy quality and analogical reasoning, and increased analogy quality after intervention. No change in reported KOC was observed, and analogy quality did not predict course performance. Findings extend understanding of strategies that can support college students' biology learning. Researchers and practitioners can leverage our approach to teaching analogies in their own research and classrooms to support students' SRL, analogical reasoning, and learning.

INTRODUCTION

Bandura (2002) contends that higher-order skills and capabilities are required to “fulfill complex occupational roles and to manage the intricate demands of contemporary life” (p. 4). It is also widely recognized that students enrolled in science, technology, engineering, and mathematics (STEM) programs, such as biology, must tap higher-order thinking skills to be successful (National Science and Technology Council, 2018). This has led many to call for an increased emphasis on teaching and supporting students' self-regulated learning (SRL; National Council of Teachers of Mathematics, 2000; Bandura, 2002; European Union Council, 2002; DiDonato, 2013). SRL is the extent to which a student deliberately plans, monitors, and regulates behavioral, motivational, and cognitive processes in pursuit of a goal (Hadwin *et al.*, 2018). Self-regulated learners engage several subprocesses. Two of the most studied include self-regulation strategies and metacognition (Pintrich, 2000), both targeted in this study. Other subprocesses not addressed in this work include motivational processes. See Panadero (2017) and Puustinen and Pulkkinen (2001) for comprehensive reviews of SRL models.

Research has shown that higher-achieving students use self-regulation strategies more frequently and more effectively than lower-achieving peers (Zimmerman, 1986, 2002; Dent and Koenka, 2016). SRL has also been closely associated with students' academic delay-of-gratification and performance calibration (Chen and Bembenuddy, 2018). Despite the myriad benefits gleaned from effectively self-regulating learning, research also finds that students rarely have the self-regulatory skills sought by

Vicente Talanquer, *Monitoring Editor*

Submitted May 10, 2022; Revised Dec 7, 2022;
Accepted Dec 20, 2022

CBE Life Sci Educ March 1, 2023 22:ar10

DOI:10.1187/cbe.22-05-0084

*Address correspondence to: Joseph C. Tise
(tise.joseph@gmail.com).

© 2023 J. C. Tise *et al.* CBE—Life Sciences Education © 2023 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 4.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/4.0>).

“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

employers, colleges, and trade schools (Winne and Jamieson-Noel, 2003; DiDonato, 2013) and required for successful careers in biology and related STEM fields. Further, recent research also indicates that students rarely regulate their learning without explicit prompting (Lazonder and Rouet, 2008; Raes *et al.*, 2016). These findings may be due in part to the contexts in which SRL is typically studied—namely, highly controlled, researcher-generated, rather than authentic learning environments. This research, taken together, has resulted in calls for instruction that promotes students' SRL within authentic learning environments (National Council of Teachers of Mathematics, 2000; European Union Council, 2002; DiDonato, 2013). The current study addressed these deficiencies and calls by implementing a contextualized learning strategies intervention in a university biology course. The intervention supported learner's generation of analogies as a learning strategy. Using a within-subjects, pre/post design, we hypothesized that students would be able to generate analogies, they would get better at generating analogies, their generated-analogy quality would positively predict analogical reasoning and final course grade, and reported knowledge of cognition (KOC; a component of metacognition; Schraw and Moshman, 1995; Tanner, 2012) would increase after intervention.

LITERATURE REVIEW

The current study tested a learning strategy intervention within a postsecondary biology course. The intervention was grounded in SRL theory, which positions strategy use and metacognition as two critical subprocesses for successful SRL (Puustinen and Pulkkinen, 2001; Panadero, 2017). There are several other subprocesses involved in successful SRL, but given the focus of this study, we only discuss these two. We direct readers to other sources to learn more about other SRL subprocesses (e.g., Zimmerman, 1986; Winne and Hadwin, 1998; Puustinen and Pulkkinen, 2001; Panadero, 2017). Metacognition comprises knowledge and regulation of cognition (ROC; Schraw and Moshman, 1995; Tanner, 2012). A student's KOC includes declarative, procedural, and conditional knowledge, which pertain to knowledge of oneself as a learner and how learning occurs, and how and when to enact strategies, procedures, and skills. ROC includes several subcomponents, such as planning, monitoring, and evaluating. These subcomponents all function to modify one's cognition (Schraw and Moshman, 1995), which we conceptualize as mental processing undertaken by learners consistent with an information-processing theoretical perspective (Atkinson and Shiffrin, 1968).

Given the multifaceted nature of SRL, interventions that target SRL may focus on one or multiple associated subcomponents, such as learning strategies (e.g., self-explanation, drawing, generating analogies; McNamara *et al.*, 2004). The current intervention focused primarily on a *cognitive* learning strategy—generating analogies. Recognizing the interconnectivity of SRL constructs, we hypothesized that the benefits from the intervention may not only impact strategy use and achievement, but may also influence students' metacognitive knowledge (KOC).

Cognitive Learning Strategies

Interventions intended to promote SRL often focus on learning strategy use, which is one component of the SRL processes. Learning strategies can be defined in several ways, but for this

study, the following definition by Zimmerman (1989) is adopted: “actions and processes directed at acquiring information or skill that involve agency, purpose, and instrumentality perceptions by the learners” (p. 329). There are several categories of strategies, including motivational, metacognitive, and cognitive strategies. The current study used a cognitive strategy, so only this type is discussed further.

Cognitive learning strategies are goal-directed, intentionally invoked, effortful procedures intended to influence the learning process (Weinstein and Mayer, 1986; Weinstein and Meyer, 1991; Dinsmore, 2018). Because strategies are consciously implemented and are controllable, students need to have sufficient knowledge about the strategies and sufficient motivation to apply them (Wittrock, 1990; Donker *et al.*, 2014). Thus, effective strategy interventions should focus on teaching students *when* to use the strategy in addition to *how* (Pressley, 2000) and should also emphasize the utility value of the strategy—defined as the “usefulness” of a task as it pertains to a learner's current or future plans (Eccles and Wigfield, 2020). Therefore, the most effective strategy instruction programs emphasize metacognitive conditional knowledge (Schraw and Gutierrez, 2015) and motivational components such as utility value and self-efficacy (Donker *et al.*, 2014).

Analogical Reasoning. Relational reasoning is “the foundational human ability to discern patterns within any stream of information” (Dumas, 2017, p. 1). Relational reasoning comprises four types of reasoning—*analogical*, *anomalous*, *antinomous*, and *antithetical*—and has been supported as a type of cognition known to positively impact preparedness of STEM professionals (Thiry *et al.*, 2011). It has further been posited that relational reasoning is a foundational cognitive ability involved in complex problem solving (Bassok *et al.*, 2012; Holyoak, 2012; Dumas, 2017), which may explain why expert and novice medical professionals appear to differentially apply relational reasoning (Dumas *et al.*, 2014). For definitions and examples of all four types of relational reasoning, please see Dumas *et al.* (2014). Although all types of relational reasoning are important, analogical reasoning is commonly employed in educational settings as an explicit learning strategy taught to students. Analogical reasoning involves uncovering structural similarities among concepts (Gentner, 1983; Alexander *et al.*, 2016; Dumas, 2017). The power of analogical reasoning can be leveraged by using analogies as a cognitive learning strategy. For example, Richland and McDonough (2010) reported students could be taught to effectively use analogies to solve mathematics problems. They found students' learning benefited when they were supported in identifying relational similarities between two combination/permutation problems. These results were replicated with problems that addressed proportional reasoning.

Although at least four forms of relational reasoning exist (Dumas *et al.*, 2013; Dumas, 2017), analogical reasoning is the most studied, perhaps due to its empirical link to student success within many domains, among them reading (Ehri *et al.*, 2009), mathematics (Richland and McDonough, 2010), chemistry (Trey and Khan, 2008), and biology (Grotzer *et al.*, 2017; Emmons *et al.*, 2018). Analogy is a powerful type of reasoning that elucidates the *deep structure* of a relationship between two concepts, even if *attributes* (surface features) of those concepts differ (Gentner, 1983). Attributes might include characteristics

such as color, shape, or size. They may cause two concepts to appear totally different, even if their deep structure is similar. The deep structure of a relationship is that which makes similar two seemingly unlike concepts. For example, a planet and an electron may seem different in terms of surface features (attributes), but they both revolve around a central entity. The latter fact is part of the deep structure of their relationship (Gentner, 1983). In this way, two entities can share a deep-structure relationship absent any attributional similarities—although they *can* share attributional similarities. Effective analogical reasoning ought to promote elaborative and organizational processing (Atkinson and Shiffrin, 1968; Mayer, 1996) by connecting new information to prior knowledge and helping the learner create, modify, and integrate coherent mental models (or schemas) for the new information (Sweller, 1988).

Analogical reasoning is malleable and teachable and can be promoted through strategy instruction (Alexander *et al.*, 1989; Richland and McDonough, 2010; Dumas, 2017). For example, Alexander *et al.* (1989) taught 132 undergraduate education majors how to analyze the component parts (Sternberg, 1977) of nonverbal, four-term analogies (e.g., A:B::C:D), visual matrix analogies (like those seen on Raven's matrices; Raven, 1958), and verbal analogies embedded within science texts. They found that students trained in science content and analogical reasoning improved their scores on several measures, including one of science analogies.

Researchers in a more recent study (Hattan, 2019) taught fifth- ($n = 78$) and sixth-grade ($n = 71$) students to use the four forms of relational reasoning while reading expository text to aid comprehension. In this intervention, students generated questions that pertained to each form of relational reasoning. For analogy, students asked how something in their own lives was similar to something in the expository text. Students trained in relational-reasoning question generation significantly outperformed a second intervention group (knowledge mobilization) and control students on a comprehension measure.

Despite findings from these initial studies, there are four main gaps in the analogies research that we have identified. First, previous important work has not adequately investigated students' ability to generate their *own* analogies after training. Instead, most of the research heretofore has focused on training students to analyze and use already-constructed analogies to learn or included traditional multiple-choice measures of analogical reasoning as an outcome.

Second, the few studies that did ask students to generate their own analogies did not analyze these generated analogies as data. Doing so could reveal important insights about how effectively (or ineffectively) students generate analogies, their developing understandings of key relationships among content, how the quality of their generated analogies relates to other individual differences, how their generated-analogy quality relates to course performance, and other important information.

Third, most of the analogies research to date has been undertaken outside authentic learning environments. Considering the established use of analogical reasoning among experts in the medical and other fields (Dumas *et al.*, 2014), it is important to study the analogies as a learning strategy within postsecondary biology courses, because they typically precede medical and other STEM-related careers in which analogical reasoning would be fruitful.

Finally, generating analogies may benefit students' metacognitive awareness—specifically KOC (Schraw and Dennison, 1994; Schraw and Moshman, 1995). When students attempt to create an analogy, they must consider what they know and do not know about the content. Prior knowledge (which helps constitute the base analogue) becomes conscious as it is drawn into working memory and mapped to the novel content analogically. Metacognitive monitoring ought to monitor the analogy-creation process, and this monitoring should result in metacognitive experiences (Flavell, 1979) of either successful mapping (and thus a positive evaluation of one's knowledge) or unsuccessful mapping (and thus a negative evaluation of one's knowledge). Such metacognitive experiences would thus inform and perhaps modify a student's KOC.

The current study targets these gaps in analogies research. In this study, students were not only taught what analogies were, when to use them, and why they were useful, but also how to generate them using their biology course content (focused on plant and animal physiology). We also collected students' generated analogies and analyzed how analogical quality changed over time. Finally, we investigated the relationships between generated-analogy quality and academic performance, as well as individual differences (e.g., KOC).

PRESENT STUDY

In the present study, we aimed to bolster our knowledge of benefits gleaned from analogies as a cognitive learning strategy by studying the effects of an analogies intervention. The intervention was implemented within a university biology course designed for second-year students and thus extends analogy intervention research into more authentic learning contexts. This helps answer the call from some researchers (e.g., Dunlosky *et al.*, 2013) to investigate the utility of analogies in representative contexts, and how analogical reasoning relates to other psychological constructs (Dumas, 2017).

Research Questions

Table 1 outlines the research questions and hypotheses for the present study. Generally, we were interested in the relations among the intervention, analogical reasoning, the quality of students' generated analogies, their reported KOC, and their final course grades.

METHOD

Undergraduate students in an introductory biology course participated in the four-part intervention. All students in the course ($n = 451$) were exposed to the intervention, as it was integrated into the course structure and learning management system (LMS; see *Intervention Overview* for further details); however, only students who volunteered to participate in the research by consenting to release their course grades and analogy data ($n = 321$) were asked to complete pre- and posttest measures. Students who did not volunteer to participate did not complete the pre- and posttest measures and did not consent to release their analogy data for the current study. Only students who had complete data on primary variables and provided consent were included for analyses. Thus, the final sample for the present study included 179 students. Absent the ability to form a control group, research questions were amended to be relational and predictive in nature rather than causal. Table 2 presents

TABLE 1. Research questions and corresponding hypotheses

Research questions	Hypotheses
1. Can students generate analogies pertaining to course content, and if so, does generated-analogy quality increase throughout the intervention?	1. Students will generate analogies pertaining to course content, and generated-analogy quality will increase throughout the intervention.
2. Does generated-analogy quality positively predict post-intervention analogical reasoning, controlling for pre-intervention analogical reasoning?	2. Generated-analogy quality will positively predict post-intervention analogical reasoning, after controlling for pre-intervention analogical reasoning.
3. Does generated-analogy quality positively predict final course grade, controlling for prior course performance?	3. Generated-analogy quality will positively predict final course grade after controlling for prior course performance.
4. Does reported KOC increase from pre- to post-intervention, and is there an interaction with generated-analogy quality?	4A. Reported KOC will increase from pre- to post-intervention. 4B. Increases in KOC will interact with generated-analogy quality.

descriptive statistics regarding participants' gender, concurrent biology course work, number of students receiving extra help, and year of study.

To confirm the students included in the final analytic sample were representative of the broader course enrollment, we conducted a series of independent-samples *t* tests comparing them with students who consented to participate but did not have complete data on primary variables (and thus were excluded). It could be argued that students who had incomplete data due to attrition from pre to post survey and/or blank analogy data ($n = 142$) were systematically different from those with complete data ($n = 179$) in ways that might affect the primary variables under study. Thus, the pre-survey values (when available) for the two groups were compared to assess any differences in self-efficacy for using analogies, self-efficacy for learning biology, KOC, use of other SRL strategies, and analogical reasoning. No significant differences were found between the two groups on any of these variables. However, those with incomplete data on all primary variables did end up with a lower final course grade (mean = 84.90%), on average, than those with complete data on all primary variables (mean = 89.88%, mean difference = -4.98%, $t_{(272,44)} = -4.48$, 95% bootstrapped CI [-7.17, -2.70], Cohen's $d = 0.51$). Standard errors and confidence intervals were corrected using 1000 bootstrap samples and degrees of freedom were adjusted for unequal variances in the final course grade analysis.

Intervention Overview

The intervention was delivered through four online activities in the course LMS. Students completed each activity outside class

TABLE 2. Demographic information for participants

Demographics ^a	<i>N</i>	%
Male	46	25.70
Female	132	73.70
Taking at least one other biology course	28	15.64
Receiving outside help for the course	45	25.14
First year	82	45.80
Second year	52	29.10
Third year	31	17.30
Fourth year	12	6.70
Fifth year+	2	1.10

^aOne student did not report their gender. Outside help may include regular office hour visits, on- or off-campus tutoring, formal study groups, or informal help from friends.

time within 5 days of release, but each activity was completed sequentially and in the same order. Each activity was assessed for completion, and those who completed all activities were awarded extra credit equal to 1% of the final course grade. The four activities collectively addressed three core components of the intervention that align with elements of the transactional strategies instruction approach (Brown *et al.*, 1996). These core components included a focus on declarative knowledge of analogies, metacognitive conditional knowledge of analogies, and practice with analogies. Activities 1 and 2 both assessed the declarative knowledge and practice components, while activities 3 and 4 addressed the metacognitive conditional knowledge and practice components. Students were instructed in class to contextualize their responses within the course content covered at the associated time of the semester. That is, generated analogies were supposed to tie to course content in some way, but were not required to *exclusively* include course content.

Activity 1 first described Wittrock's (1994) model of generative learning and how analogies can aid in generating the two types of meaningful relationships outlined in the model of generative learning (i.e., relations among to-be-learned information and relations between to-be-learned information and prior knowledge). It also described how to create effective analogies by defining "analogies," "deep structure," and "surface features" and how these concepts relate to form an analogy. An example was also given in which the relationships and surface features were explicated. Supplemental Appendix A includes a copy of activity 1. Activity 2 reiterated this information, introduced some metacognitive conditional knowledge associated with when to use analogies, and provided new practice opportunities. Activity 3 addressed metacognitive conditional knowledge pertaining to analogies in more detail. Activity 4 reiterated this information. For the practice component, each activity included four unique questions created by M.S.D. and J.C.T. that demonstrated analogies and analogical reasoning within the course content covered in the corresponding unit. Students answered each of these four example analogy questions, and then in a fifth question (for activities 1 and 2; fifth and sixth question for activities 3 and 4), were asked to generate their own analogy with course content. An example question from activity 4 is: "How is the linear flow of electrons between Photosystems II and I similar to water moving past a water wheel in a mill?" Activities 1 and 2 included one question that scaffolded students to generate their own analogy and answer it. Activities 3 and 4 included two such questions. An example question of this type is:

Now you try. Generate and attempt to answer your own analogy. A basic template to help you start follows: How is _____ [new concept] related to what you already know about _____ [experience with something similar to this concept]?

Measures

Two primary constructs (analogical reasoning and KOC) were measured via identical pre- and postassessments. These assessments included the Verbal Test of Relational Reasoning (vTORR; Alexander *et al.*, 2016), the knowledge-of-cognition subscale of the 19-item version (Harrison and Vallin, 2018) of the Metacognitive Awareness Inventory (MAI; Schraw and Dennison, 1994), and demographic questions. Course performance data and intervention responses were also collected from the course LMS, and intervention responses were coded for analogical complexity (see Supplemental Appendix B). Two students at time 3 and five at time 4 did not provide a “first” analogy (i.e., response to question 5), but did provide a “second” analogy (i.e., response to question 6). In these cases, we used students’ “second” analogies in place of their missing “first” analogies. Thus, all students had complete analogy data at each time point.

vTORR. Analogical reasoning was measured by the vTORR (Alexander *et al.*, 2016). The vTORR assesses four types of relational reasoning: analogical, anomalous, antinomous, and antithetical. Only the analogical reasoning subscale was used in the current study. Participants were presented two practice questions followed by eight questions used in score calculation. Each question presented a relationship and asked the participant to choose one of four answer choices that demonstrated a similar relationship. An example question stem is: “The inspired author opened the computer and began pouring words onto the page.” The provided answer choices (correct choice marked with *) for this stem were: a) The excited composer turned off the record and began to write a melody, b) The cheerful artist uncovered the easel and envisioned his composition, c) The exhilarated actor opened her script and recited her lines, or *d) The motivated sculptor picked up his chisel and started creating a statue. Test–retest reliability, latent factor reliability (coefficient H), and three sources of validity evidence (convergent, discriminant, and internal structure) have demonstrated the vTORR to be psychometrically sound for use at a single time point or multiple time points (Alexander *et al.*, 2016). Importantly, Alexander *et al.* (2016) demonstrated that scores on the vTORR were not accounted for simply by linguistic ability (measured by the vocabulary cloze task from the Graduate Record Examination). The analogies subscale demonstrated somewhat lower reliability than desired in the present sample at pre survey ($\alpha = 0.60$) and post survey ($\alpha = 0.61$), but was still above the recommended threshold of 0.50 for use in research (Evers, 2001). Reliability estimates were also higher than observed values in prior research (Kottmeyer *et al.*, 2019). Such relatively low reliability is likely due partially to the dichotomous scoring of these multiple-choice items, because Cronbach’s alpha is calculated by correlating each item value (in this case, 0 = incorrect or 1 = correct) with the sum of the other item values and then comparing those values with the total variance observed across all items (Cronbach, 1951). Thus, the length of the scale and how each item is scored (e.g., dichotomous,

polytomous, continuous) can influence Cronbach alpha values, because these factors limit the total possible variance in the measure. Still, a confirmatory factor analysis revealed the eight items loaded onto a single factor, as originally intended, with good model fit at both pre survey ($\chi^2 = 34.90$, $p = 0.021$, comparative fit index [CFI] = 0.97, root-mean-square error of approximation [RMSEA] = 0.05, standardized root-mean-square residual [SRMR] = 0.06) and post survey ($\chi^2 = 36.20$, $p = 0.015$, CFI = 0.95, RMSEA = 0.05, SRMR = 0.07).

KOC. A shortened version of the MAI was used to measure students’ reported KOC (Schraw and Dennison, 1994). The original measure contained 17 KOC items, but a recent study suggests that a shortened version holds a better factor structure (Harrison and Vallin, 2018). Students were thus asked to rate each of eight statements from 1 (not at all typical of me) to 5 (very typical of me). For the short version, the maximum-likelihood estimation reliability estimate for the KOC subscale was 0.80 (Harrison and Vallin, 2018), indicating sound reliability. An example item from the KOC subscale is: “I know what kind of information is most important to learn.” Internal consistency reliability estimates for the current sample were strong for both pre ($\alpha = 0.85$) and post surveys ($\alpha = 0.86$).

Coded Analogies. We began the coding process by examining and coding all analogies submitted by students who had provided consent to participate in the study. Students’ generated-analogy responses ($n = 2,226$ across all students in the course) were coded from 0 to 5 to assess the level of processing required of the student to produce the response. After completely blank analogy responses ($n = 551$) were discarded, a corpus of non-blank analogies ($n = 1675$) remained and were coded. Next, we removed students from the study if they had incomplete data on primary variables (i.e., generated analogies, course performance data, and survey data for analogical reasoning and KOC). Due to missingness on these primary variables, the final analytic sample comprised 179 students, and thus the number of coded analogies included for analyses was 716—only students’ first responses for activities 3 and 4 were included (i.e., four analogies per student were included in analyses). The decision to include students’ first responses for activities 3 and 4 is discussed further in the *Procedure* section. The coding framework is provided in Supplemental Appendix B.

Final Course Grades. Final course grades (expressed in percent) were retrieved from the LMS after all grades for the semester were entered. The final course grade calculation comprised daily clicker questions (6%), a laboratory component (25%), four regular exams (52%), and one final exam (17%).

Procedure

The intervention was delivered through four activities distributed and collected through the course LMS. Students submitted responses to each activity within 5 days of its release—submissions after the fifth day were excluded from analyses. In total, 178 students were excluded from analyses because they did not submit all analogies on time or did not submit all analogies. Three weeks after the final activity was released, students responded to the postassessments.

TABLE 3. Descriptive statistics for primary variables

Measure	Mean (SD)	Median	Min.–Max.
Exam 1 percent	87.12 (10.04)	88.46	40.00–100.00
Analogical reasoning	5.72 (1.77)	6.00	1–8
KOC	30.94 (4.69)	31.00	13.00–40.00
Analogy 1	2.73 (1.61)	3.00	0–5
Analogy 2	2.98 (1.57)	4.00	0–5
Analogy 3	3.08 (1.53)	4.00	0–5
Analogy 4	3.15 (1.49)	4.00	1–5
Analogies (total)	11.93 (4.56)	13.00	2–19
Post analogical reasoning	5.78 (1.83)	6.00	0–8
Post KOC	30.79 (4.79)	31.00	17.00–40.00
Final course grade percent	89.91 (8.79)	91.79	50.40–102.75

After data collection was complete, the authors began analyzing the qualitative data. A coding framework to assess students' generated analogies (question 5 on activities 1 and 2, and questions 5 and 6 on activities 3 and 4) was then created. For activities 3 and 4, the student's first analogy score (i.e., question 5 in the activities) was used in analyses to facilitate better comparisons with those generated in activities 1 and 2 (in which only one analogy was generated).

To develop the coding framework, 100 student-generated analogies (25 from each activity) were randomly selected. J.C.T. and R.A.S. then sorted each response into emergent categories based on similarity. These categories were then examined for distinctive features, which became the basis for the coding framework. M.S.D. (an expert in biology) was then consulted to ensure its applicability to the biology course content. After confirmation from M.S.D., the framework was iteratively refined through application to new batches of randomly selected responses (12 from each activity per batch). J.C.T., R.A.S., and T.M.Y. discussed each round of coding to identify coding discrepancies and confusion in the framework. After four rounds of refinement, the final framework was developed and satisfactory interrater reliability was achieved (agreement = 81%; Miles and Huberman, 1994; Saldaña, 2013).

Using the developed framework, J.C.T. and T.M.Y. independently coded batches of analogies. To ensure ongoing interrater reliability, 20% of the analogies were double coded, and discrepancies were resolved through consensus coding. Our decision to double code 20% of the analogies was guided by recommendations from the literature (Creswell and Miller, 2000; Campbell et al., 2013). The average interrater agreement of this 20% segment was 76%, and the Spearman's rho correlation was 0.91. Although disagreement was present, the magnitude of such disagreement was low. The high correlation between the two coders partially ameliorates the slightly lower than expected agreement. Bias was mitigated by deidentifying the analogies during coding, such that the raters could not know which participant created the analogy or which activity (and thus part of the intervention) the analogy came from. A total of 1607 student-generated analogies were coded.

RESULTS

Descriptive statistics for all relevant variables are presented in Table 3; Table 4 presents the correlations among these variables.

Research Question 1: Change in Generated-Analogy Quality

The first research question investigated whether students could generate their own analogies pertaining to course content, and if so, whether students' generated-analogy quality increased throughout the intervention. We hypothesized that students would be able to generate analogies using course content and that the quality of generated analogies would increase over the intervention. Based on the median scores of generated analogies at each time point (see Table 3) it was clear that at least half of the students could generate analogies using course content at time points 2–4. Fifteen (2.09%) analogy responses across all time points were coded "0," indicating failure to even approximate an analogy. And the results of an omnibus repeated-measures ANOVA (with Greenhouse-Geisser-corrected degrees of freedom) indicated the quality of students' generated analogies was not equal at all time points; $F(2.84, 504.99) = 4.09, p = 0.008, \eta_p^2 = 0.022$.¹ This significant F statistic indicates that variance in analogy scores is attributable to time point, and the associated effect size is small (Cohen, 1988). Post hoc analyses with a Bonferroni adjustment for family-wise error indicated that participants' generated analogies at time 4 (i.e., activity 4) were, on average, rated higher than the generated analogy at time 1 (mean difference = 0.42, $p = 0.02, d = 0.26$). Further, participants' generated analogy at time 3 was *almost*, on average, rated higher than the generated analogy at time 1 (mean difference = 0.35, $p = 0.058, d = 0.19$). Effect sizes (Cohen's d) associated with both of these pairwise comparisons indicated a small effect (Cohen, 1988). No other pairwise comparisons were statistically significant. Figure 1 displays these results, and the statistically significant linear trend, $F(1, 178) = 9.32, p = 0.003, \eta_p^2 = 0.05$, indicating a positive change from time 1 to time 4 can also be seen. To illustrate this positive change, one selected student's responses from times 1 and 4 are reproduced below.

Time 1 response: "Developmental processes in the embryo occur in a head to tail manner the same way a child grows more so in height than in width." (Coded 1)

Time 4 response: "Digestion of food is like making mulch. First the tree must be cut down into logs (mechanical digestion) and then ground into much smaller pieces for its use as mulch (chemical digestion)." (Coded 4)

¹The Greenhouse-Geisser correction was applied to this test given a significant test of sphericity.

TABLE 4. Correlations among primary variables

	1	2	3	4 ^a	5 ^a	6 ^a	7 ^a	8	9	10
1. Exam 1 percent										
2. Analogical reasoning	0.22**									
3. KOC	0.18*	0.08								
4. Analogy 1 ^a	0.17*	0.17*	0.04							
5. Analogy 2 ^a	0.19*	0.14	0.12	0.31**						
6. Analogy 3 ^a	0.17*	0.10	0.09	0.31**	0.47**					
7. Analogy 4 ^a	0.11	0.19*	0.02	0.25**	0.39**	0.45**				
8. Analogies overall	0.28**	0.23**	0.13	0.67**	0.74**	0.74**	0.70**			
9. Post analogical reasoning	0.13	0.50**	0.10	0.16*	0.18*	0.22**	0.28**	0.33**		
10. Post KOC	0.34**	0.05	0.59**	0.06	0.19*	0.17*	0.13	0.22**	0.15*	
11. Final course grade percent	0.82**	0.20**	0.23**	0.19*	0.23**	0.18*	0.18*	0.30**	0.11	0.40**

^aCorrelations with these variables are Spearman’s rho coefficients because of their ordinal scale.

* $p < 0.05$.

** $p < 0.01$.

In sum, findings not only support that students could generate analogies related to course content, but also that students’ generated-analogy quality improved throughout the intervention.

Research Question 2: Analogy Quality and Analogical Reasoning

The second research question addressed whether students’ overall generated-analogy quality positively predicted post-intervention analogical reasoning, after controlling for pre-inter-

vention analogical reasoning. We hypothesized that it would, and this hypothesis was supported through multiple linear regression analysis with Bonferroni-adjusted significance testing; $F(2, 176) = 33.29, p < 0.001, R^2 = 0.30$. Overall generated-analogy quality was a significant predictor (standardized $B = 0.22, t = 3.43, p = 0.002$) of post-intervention analogical reasoning, above and beyond pre-intervention analogical reasoning, which was also a significant predictor (standardized $B = 0.45, t = 6.93, p < 0.001$). Overall generated-analogy quality accounted for an additional 4.70% of the variance in students’

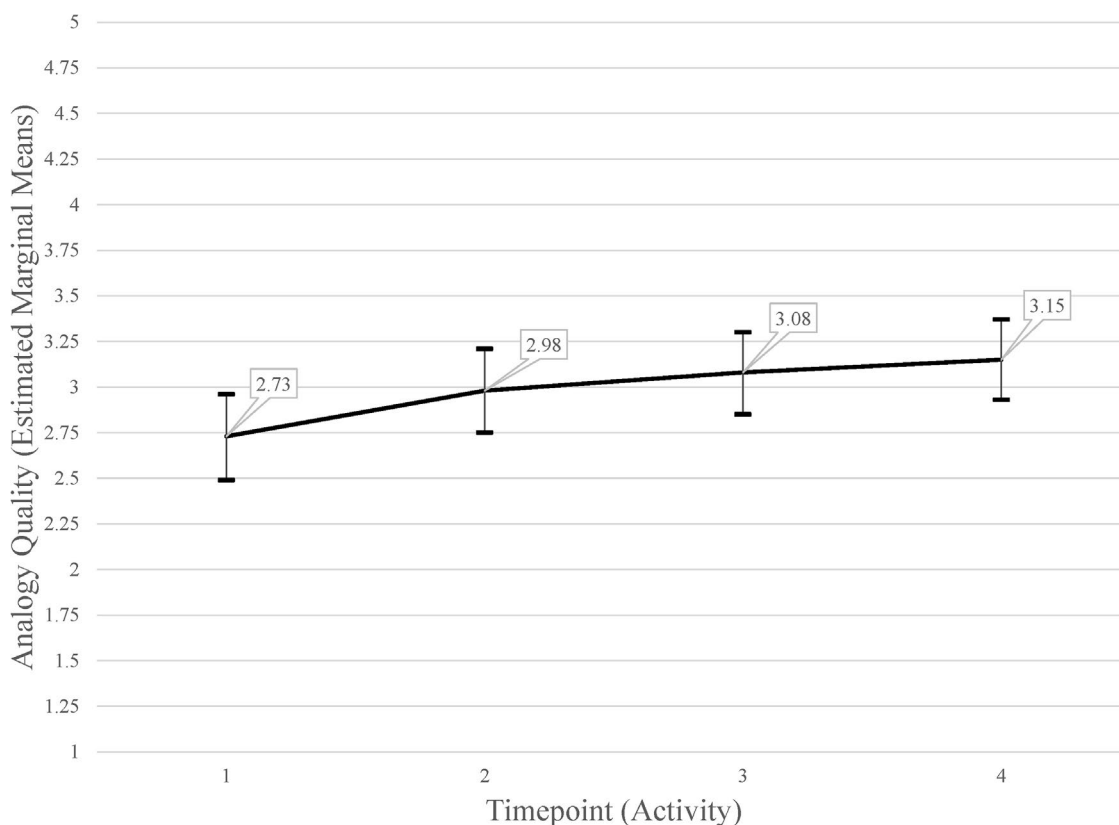


FIGURE 1. Estimated marginal means of students’ generated-analogy quality at each intervention time point.

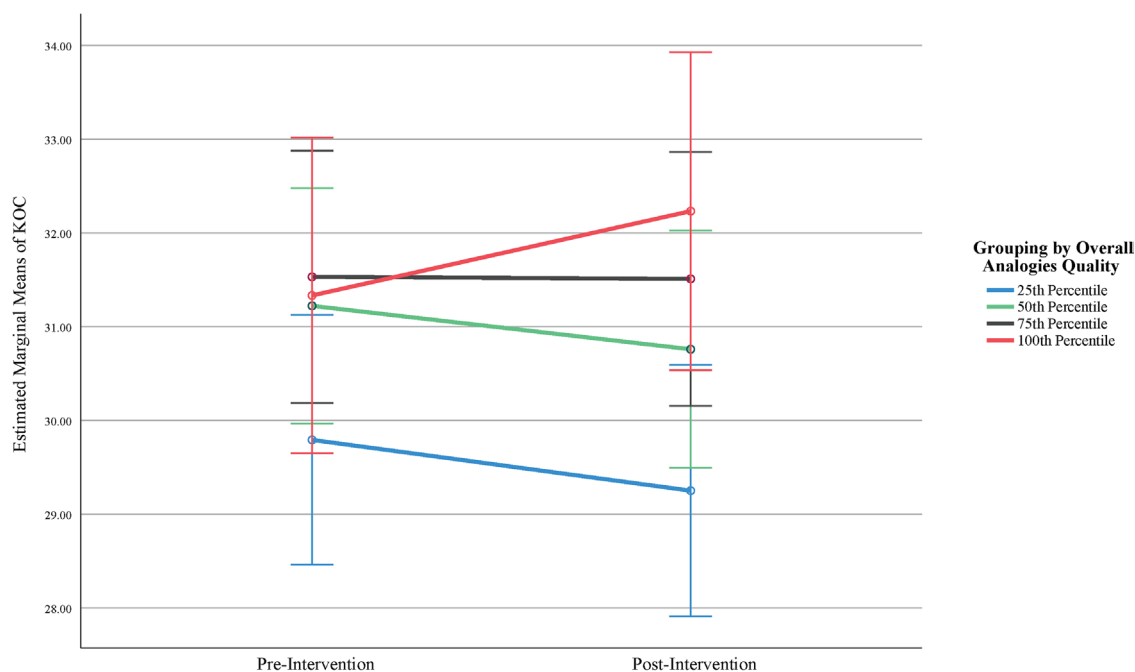


FIGURE 2. Estimated marginal means of KOC scores at pre- and post-intervention, grouped by quartiles of students' overall generated analogies scores.

post-intervention analogical reasoning scores beyond pre-intervention analogical reasoning scores (R^2 change = 0.047, $p < 0.001$). Thus, even after controlling for pre-intervention analogical reasoning, students' generated-analogy quality positively predicted post-intervention analogical reasoning.

Research Question 3: Analogy Quality and Course Performance

The third research question sought to determine whether overall generated-analogy quality positively predicted final course grade after controlling for prior course performance (exam 1 grade). We hypothesized that it would, because students able to create higher-quality analogies theoretically have a deeper understanding of the course content, which should translate to a higher course grade. Multiple linear regression analysis with Bonferroni-adjusted significance testing did not support this hypothesis. Only exam 1 (standardized $B = 0.82$, $t = 19.29$, $p < 0.001$) statistically significantly predicted final course grade; $F(1, 177) = 372.21$, $p < 0.001$, $R^2 = 0.68$. Overall generated-analogy quality was not a significant predictor of final course grade (standardized $B = 0.07$, $t = 1.60$, $p = 0.22$). To further explore this null result, we calculated change in analogy scores (from time 1 to time 4) to see whether this change score predicted course performance. Regression analysis revealed this change score also did not predict final course grade above and beyond exam 1 grades (standardized $B = 0.03$, $t = 0.63$, $p = 0.53$). Regarding research question 3, results showed that generated-analogy quality did not predict final course grade above and beyond prior course performance.

Research Question 4: Analogy Quality and KOC

The fourth research question investigated whether reported KOC increased from pre- to post-intervention, and if so, whether

there was an interaction with generated-analogy quality. We hypothesized that KOC would increase from pre- to post-intervention and that the increase would be more pronounced for students with higher overall generated-analogy quality. Results from a 2×4 mixed-design ANOVA with Bonferroni corrections for family-wise error did not support this hypothesis. Students were split into four groups based on their percentile score for their total generated analogies score (25th, 50th, 75th, and 100th percentiles). These four groups were the independent variable in a 2×4 mixed-design ANOVA. Results indicated that students' average reported KOC did not increase from pre- to post-intervention; there was no main effect for time; $F(1, 175) = 0.009$, $p = 0.92$, $\eta_p^2 = 0.000$. Further, there was no main effect for group, $F(3, 175) = 2.54$, $p = 0.06$, $\eta_p^2 = 0.042$; and no time \times group interaction, $F(3, 175) = 0.85$, $p = 0.47$, $\eta_p^2 = 0.014$. Figure 2 displays these results.

To further explore the potential relationship between students' generated-analogy quality and their KOC, we posed a post hoc, slightly modified version of the fourth research question that mimicked the format of research questions 2 and 3. Specifically, we used multiple regression to determine whether generated-analogy quality predicted post-intervention KOC after controlling for pre-intervention KOC. Results from this analysis (with Bonferroni-adjusted significance testing) indicated that students' overall analogy quality *did* predict post-intervention KOC, above and beyond pre-intervention KOC; $F(2, 176) = 52.12$, $p < 0.001$, $R^2 = 0.37$. Specifically, overall analogy quality (standardized $B = 0.14$, $t = 2.35$, $p = 0.04$) accounted for an additional 2% ($p = 0.02$) of the variance in post-intervention KOC beyond pre-intervention KOC (standardized $B = 0.57$, $t = 9.53$, $p < 0.001$).

To summarize, regarding our fourth research question, results indicated that students' KOC did not increase from

pre- to post-intervention, and thus, there was no interaction with generated-analogy quality. However, exploratory regression analysis showed that analogy quality predicted post-intervention KOC above and beyond pre-intervention KOC and explained an additional 2% of the variance in post-intervention KOC.

DISCUSSION

This study addressed four major gaps in the literature. First, it was one of the first to investigate whether students could generate their own analogies, and if so, whether they could get better at doing so. We found that students could generate contextualized analogies and improve the analogical complexity of their analogies over time. Related to the first gap, the second gap this study helped address was the dearth of empirical work that incorporated student-generated analogies as data. No coding framework existed (before our study) to code generated analogies for analogical complexity based on Gentner's structure-mapping theory of analogy (Gentner, 1983). Future researchers can now leverage our developed coding scheme (Supplemental Appendix B) as a tool to further investigate the quality of analogies. The coding framework can surely be refined, but serves as an initial tool. Third, this study is one of few to investigate analogical reasoning within a representative educational context (i.e., a postsecondary biology course). Our data show that generating analogies contextualized to an authentic biology course is possible and predicts students' analogical reasoning and KOC. This latter relationship speaks to the fourth major gap in the literature addressed by this study. Researchers have called for analogical reasoning to be investigated in relation to other individual-difference variables (Dumas *et al.*, 2013), and our data showed that, although students' KOC did not change from pre- to post-intervention, generated-analogy quality did positively predict post-intervention KOC above and beyond pre-intervention KOC. Detailed discussions of the results pertaining to each research question follow.

Students' generated-analogy quality increased throughout the intervention, possibly because of the increased exposure to and practice with content-based analogies. One might argue the observed small effect size was caused by maturation, such that learners got better at generating analogies because they had become more and more familiar with the course content. This explanation is unlikely, however, because students generated analogies using new (recently covered) content at each activity. Thus, little maturation could have occurred with each new content unit. One might also point to a testing effect (i.e., students got better at analogies because they had repeated practice with them) to explain the effect size, and in a sense, a testing effect was exactly what we targeted in the intervention. Indeed, the intention was to teach students to generate analogies and to get better at doing so. Repeated practice with analogies was in fact one of the core components of the intervention. The primary intervention goal appears to have been reached.

Analogies created at time 4 were higher quality than those created at time 1. Although students did not receive individual feedback on their generated analogies, they were able to compare their work with the instructor-generated analogies. It is likely that the more exemplar analogies they interacted with and the more practice they had generating their own analogies, the better they became at generating higher-quality analogies.

This finding is encouraging evidence that one goal of the intervention was met. Hopefully, these students continued to practice creating analogies with other biology content and were able to transfer the learning strategy to other courses. Another possible explanation for these findings is that the focus on metacognitive conditional knowledge (at times 3 and 4) was more impactful on students' generated-analogy quality compared with a focus on declarative and procedural knowledge (at time 1). While this explanation remains possible, we argue that, in fact, a combination of declarative, procedural, and metacognitive conditional knowledge supported by ample practice was necessary before any potential benefit of the intervention program could manifest. Thus, we view the increased performance at time 4 as a result of the combined effects of the intervention as a whole, rather than reflecting the impact of individual components. Such an explanation is also supported by prior strategy instruction research (Brown *et al.*, 1996; Dignath and Büttner, 2008; Donker *et al.*, 2014). Still, future research should investigate the efficacy of analogies interventions that focus exclusively on one of these intervention components to determine whether a pared-down version of this intervention may yield equivalent or even stronger effects. Although this study helped push research on analogies into more-authentic learning environments, future research should assess delayed effects of this and similar analogies interventions to determine whether students maintain and/or transfer their newly learned strategy. Maintained use and quality of generated analogies are understudied outcomes of analogical reasoning research and of strategies research more broadly. In the short-term, these results support prior research (Alexander *et al.*, 1989; Richland and McDonough, 2010; Dumas, 2017) that has found that students' analogical reasoning and ability to generate their own analogies is malleable.

On a related note, generated-analogy quality positively predicted analogical reasoning as measured by the vTORR. While the observed effect size ($\Delta R^2 = 0.047$) was smaller than we had hoped, it was statistically significant and shows that an out-of-class, assignment-based intervention can help explain change in students' analogical reasoning, even under strict statistical controls. This positive predictive relationship is encouraging and provides further evidence of validity for the measure (based on test-criterion relationships; American Educational Research Association *et al.*, 2014). Theoretically, students better able to create analogies should reason analogically more effectively. This relationship appears to have been captured in the current data. While we are unable to determine whether the intervention *caused* the increase in analogical reasoning, controlling for pre-intervention analogical reasoning lends support for potential causation. Further, if the intervention did cause an increase in analogical reasoning, this increase in analogical reasoning likely helped cause an increase in learning and, thus, course performance. A carefully designed future study could investigate this mediation hypothesis. We would expect such a relationship to manifest, given the established relationship between strategy use and learning and performance (Dent and Koenka, 2016).

Students who generate high-quality analogies should create more and deeper connections with the course content, and thus learn the material more effectively. This contention was not supported by the current data; course grades were not predicted by

the quality of generated analogies, after controlling for prior course performance. The final grade is composed of many components, such as exams, clicker points (daily warm-up questions), and an applied laboratory section. Each of these learning activities require different cognitive processing (Mayer, 1996; Kiewra, 2005), skills, and knowledge. Given the myriad demands students faced in the course, it could be that one assignment-based analogies intervention was not powerful enough to explain statistically significant variance in the final course grade. To further uncover potential relationships between generated-analogy quality and learning, future research could employ measures of learning that are more closely aligned with the cognitive processing theoretically invoked by analogical reasoning (e.g., elaborative and organizational processing).

Finally, students' reported KOC was no different after the intervention, counter to our hypothesis. We anticipated that repeated engagement with analogies would compel students to think deeply about their understanding of the content and thus guide them to develop their KOC. We also expected KOC to increase because of the conditional knowledge of analogies conveyed during the intervention (i.e., students would learn *when* to use the strategy, which would beget increased KOC). Results indicated that reported KOC did not increase, and students' analogy quality was irrelevant to this finding. Perhaps students require more practice with analogies and in more varied settings, feedback about their analogies, and/or more explicit metacognitive training to increase KOC. Or perhaps training one cognitive learning strategy is insufficient to noticeably affect KOC. This latter explanation seems in line with previous research supporting the general monitoring skill hypothesis (Schraw et al., 1995). This hypothesis suggests that metacognitive monitoring depends, in part, on a learner's domain-general metacognitive knowledge (Schraw and Nietfeld, 1998). Such domain-general metacognitive knowledge would imply that aspects of metacognition can be conceptualized as general, stable traits rather than unstable (and therefore easily malleable) contextualized events. The method employed to measure KOC could also have impacted our results. Self-report questionnaires are but one method of many that can be used to study metacognition (Schraw, 2009). It could be that the measure employed in this study was not sensitive or contextualized enough to detect any potential change in students' KOC after intervention. Finally, these null results could have manifested because of insufficient instruction on when to use analogies. Perhaps additional, more detailed instruction is required on that front to produce changes in students' KOC. Finally, although no difference from pre- to post-intervention was observed in KOC scores, a multiple regression analysis showed that generated-analogy quality did predict post-intervention KOC above and beyond pre-intervention KOC. This result provides some empirical evidence of our hypothesized link between generating analogies and invoking KOC, but more research is needed to further disentangle this relationship. We still do not know, for example, if the intervention caused students to generate higher-quality analogies, which in turn compelled them to engage and assess their KOC, or if their existing (pre-intervention) tendencies to engage and assess KOC benefited their generated-analogy quality. This distinction would be an important contribution to both metacognitive and analogical reasoning theory. Future studies should explore how interven-

tion factors might impact metacognitive awareness broadly and KOC specifically.

The results of this study reveal a promising confirmatory picture: It appears possible to teach a cognitive learning strategy to students in a real university course. The study also shed light on the relationship between analogy quality and KOC—a response to the call to investigate analogical reasoning as it relates to other individual-difference constructs (Dumas, 2017).

Taken together, the findings of the current study suggest that it is possible to teach analogies as a cognitive learning strategy to students within their normal courses of study and without any intensive materials or significant time investment (each activity only took students about 5–10 minutes to complete, on average). Future research should extend and enhance similar interventions in at least four ways. First, future interventions should use a true quasi-experimental or experimental design. Doing so will simultaneously preserve ecological validity and bolster internal validity of the study. Second, future interventions should vary the dosage and intensity of the intervention to help determine optimal intervention characteristics (e.g., de Boer et al., 2014; Schraw and Gutierrez, 2015). Third, future research should use delayed assessments of strategy use and quality of strategy use. The gains observed in analogy quality were demonstrated within a relatively short time frame (about 2 months) and were observed during intervention. It would be important to know if these gains were sustained, increased, or decreased after 2 weeks, 1 month, or even 3 months post-intervention. Finally, though analogical reasoning is touted as an important cognitive process invoked in many domains and topics (Alexander and Kulikowich, 1991; Alexander, 2019), future research should explicitly study the effectiveness of generating analogies in different domains, perhaps using a within-subjects design. It is possible that generating analogies is more beneficial for some academic domains. Teasing out the domain-specificity of this cognitive learning strategy would benefit researchers and practitioners in their efforts to promote strategy use.

CONCLUSIONS

In the current study, students in an authentic STEM classroom benefited after an analogies intervention. As practitioners and researchers respond to calls to increase achievement and retention within STEM courses, they can leverage these results to promote learning and analogical reasoning within authentic learning environments. Students can learn to generate analogies using course content, and they can improve the quality of their generated analogies over time. The current work also begins to uncover the relationships among analogical reasoning and other constructs. Although analogical reasoning was unrelated to KOC in these data, we encourage future researchers to re-examine this relationship with varied methods and measures.

ACKNOWLEDGMENTS

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

- Alexander, P. A. (2019). Individual differences in college-age learners: The importance of relational reasoning for learning and assessment in higher education. *British Journal of Educational Psychology*, 89(3), 416–428. <https://doi.org/10.1111/bjep.12264>

- Alexander, P. A., & Kulikowich, J. M. (1991). Domain knowledge and analogic reasoning ability as predictors of expository text comprehension. *Journal of Reading Behavior*, 23(2), 165–190. <https://doi.org/10.1080/10862969109547735>
- Alexander, P. A., Pate, P. E., Kulikowich, J. M., Farrell, D. M., & Wright, N. L. (1989). Domain-specific and strategic knowledge: Effects of training on students of differing ages or competence levels. *Learning and Individual Differences*, 1(3), 283–325. [https://doi.org/10.1016/1041-6080\(89\)90014-9](https://doi.org/10.1016/1041-6080(89)90014-9)
- Alexander, P. A., Singer, L. M., Jablansky, S., & Hattan, C. (2016). Relational reasoning in word and in figure. *Journal of Educational Psychology*, 108(8), 1140–1152. <https://doi.org/10.1037/edu0000110>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In Spence K. W., & Spence J. T. (Eds.), *The psychology of learning and motivation: Advances in research and theory* (pp. 89–195). New York, NY: Academic Press.
- Bandura, A. (2002). Growing primacy of human agency in adaptation and change in the electronic era. *European Psychologist*, 7(1), 2–16. <https://doi.org/10.1027//1016-9040.7.1.2>
- Bassok, M., Dunbar, K. N., & Holyoak, K. J. (2012). Introduction to the special section on the neural substrate of analogical reasoning and metaphor comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(2), 261–263. <https://doi.org/10.1037/a0026043>
- Brown, R., Pressley, M., van Meter, P., & Schuder, T. (1996). A quasi-experimental validation of transactional strategies instruction with low-achieving second-grade readers. *Journal of Educational Psychology*, 88(1), 18–37. <https://doi.org/10.1037/0022-0663.88.1.18>
- Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding in-depth semistructured interviews: Problems of unitization and inter-coder reliability and agreement. *Sociological Methods & Research*, 42(3), 294–320. <https://doi.org/10.1177/0049124113500475>
- Chen, P. P., & Bembenuity, H. (2018). Calibration of performance and academic delay of gratification: Individual and group differences in self-regulation of learning. In Schunk D. H., & Greene J. A. (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 407–420). New York, NY: Routledge.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124–130. https://doi.org/10.1207/s15430421tip3903_2
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- de Boer, H., Donker, A. S., & van der Werf, M. P. C. (2014). Effects of the attributes of educational interventions on students' academic performance: A meta-analysis. *Review of Educational Research*, 84(4), 509–545. <https://doi.org/10.3102/0034654314540006>
- Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review*, 28, 425–474. <https://doi.org/10.1007/s10648-015-9320-8>
- DiDonato, N. C. (2013). Effective self- and co-regulation in collaborative learning groups: An analysis of how students regulate problem solving of authentic interdisciplinary tasks. *Instructional Science*, 41(1), 25–47. <https://doi.org/10.1007/s11251-012-9206-9>
- Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, 3(3), 231–264. <https://doi.org/10.1007/s11409-008-9029-x>
- Dinsmore, D. L. (2018). *Strategic processing in education* (1st ed.). New York, NY: Routledge.
- Donker, A. S., de Boer, H., Kostons, D., Dignath van Ewijk, C. C., & van der Werf, M. P. C. (2014). Effectiveness of learning strategy instruction on academic performance: A meta-analysis. *Educational Research Review*, 11, 1–26. <https://doi.org/10.1016/j.edurev.2013.11.002>
- Dumas, D. (2017). Relational reasoning in science, medicine, and engineering. *Educational Psychology Review*, 29(1), 73–95. <https://doi.org/10.1007/s10648-016-9370-6>
- Dumas, D., Alexander, P. A., Baker, L. M., Jablansky, S., & Dunbar, K. N. (2014). Relational reasoning in medical education: Patterns in discourse and diagnosis. *Journal of Educational Psychology*, 106, 1021–1035.
- Dumas, D., Alexander, P. A., & Grossnickle, E. M. (2013). Relational reasoning and its manifestations in the educational context: A systematic review of the literature. *Educational Psychology Review*, 25, 391–427. <https://doi.org/10.1007/s10648-013-9224-4>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and socio-cultural perspective on motivation. *Contemporary Educational Psychology*, 61, 1–13. <https://doi.org/10.1016/j.cedpsych.2020.101859>
- Ehri, L. C., Satlow, E., & Gaskins, I. (2009). Grapho-phonemic enrichment strengthens keyword analogy instruction for struggling young readers. *Reading & Writing Quarterly*, 25, 162–191. <https://doi.org/10.1080/10573560802683549>
- Emmons, N., Lees, K., & Kelemen, D. (2018). Young children's near and far transfer of the basic theory of natural selection: An analogical storybook intervention. *Journal of Research in Science Teaching*, 55(3), 321–347. <https://doi.org/10.1002/tea.21421>
- European Union Council. (2002). Council Resolution of 27 June 2002 on lifelong learning. *Official Journal*. <https://op.europa.eu/en/publication-detail/-/publication/0bf0f197-5b35-4a97-9612-19674583cb5b>
- Evers, A. (2001). The revised Dutch rating system for test quality. *International Journal of Testing*, 1(2), 155–182. https://doi.org/10.1207/S15327574IJT0102_4
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906–911. <https://doi.org/10.1037/0003-066x.34.10.906>
- Gentner, D. (1983). Structure mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155–170.
- Grotzer, T. A., Solis, S. L., Tutwiler, M. S., & Cuzzolino, M. P. (2017). A study of students' reasoning about probabilistic causality: Implications for understanding complex systems and for instructional design. *Instructional Science*, 45(1), 25–52. <https://doi.org/http://dx.doi.org/10.1007/s11251-016-9389-6>
- Hadwin, A. F., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In Schunk D. H. & Greene J. A. (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 83–106). New York, NY: Routledge.
- Harrison, G. M., & Vallin, L. M. (2018). Evaluating the metacognitive awareness inventory using empirical factor-structure evidence. *Metacognition and Learning*, 13(1), 15–38. <https://doi.org/10.1007/s11409-017-9176-z>
- Hattan, C. (2019). Prompting rural students' use of background knowledge and experience to support comprehension of unfamiliar content. *Reading Research Quarterly*, 54(4), 451–455. <https://doi.org/10.1002/rrq.270>
- Holyoak, K. J. (2012). Analogy and relational reasoning. In Holyoak K. J., & Morrison R. G. (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 234–259). Oxford, UK: Oxford University Press.
- Kiewra, K. A. (2005). *Learn how to study and SOAR to success*. Upper Saddle River, NJ: Pearson Education.
- Kottmeyer, A. M., van Meter, P. N., & Cameron, C. E. (2019). The role of representational system in relational reasoning. *Journal of Educational Psychology*, 112(2). <https://doi.org/10.1037/edu0000374>
- Lazonder, A. W., & Rouet, J.-F. (2008). Information problem solving instruction: Some cognitive and metacognitive issues. *Computers in Human Behavior*, 24(3), 753–765. <https://doi.org/10.1016/J.CHB.2007.01.025>
- Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review*, 8(4), 357–371. <https://doi.org/10.1007/BF01463939>
- McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research*

- Methods, Instruments, and Computers*, 36(2), 222–233. <https://doi.org/10.3758/BF03195567>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: A sourcebook of new methods* (2nd ed.). Thousand Oaks, CA: Sage.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics* (Vol. 1). Reston, VA.
- National Science and Technology Council. (2018). *Charting a course for success: America's strategy for STEM education*. Washington, DC: U.S. Government Office of Science and Technology.
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8(April), 1–28. <https://doi.org/10.3389/fpsyg.2017.00422>
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In Boekaerts, M., Pintrich, P. R., & Zeidner, M., (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.
- Pressley, M. (2000). What should comprehension instruction be the instruction of? In Pearson, P. D., Barr, R., & Kamil, M. L. (Eds.), *Handbook of reading research* (pp. 546–561). Mahwah, NJ: Erlbaum.
- Puustinen, M., & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research*, 45(3), 269–286. <https://doi.org/10.1080/00313830120074206>
- Raes, A., Schellens, T., De Wever, B., & Benoit, D. F. (2016). Promoting metacognitive regulation through collaborative problem solving on the Web: When scripting does not work. *Computers in Human Behavior*, 58, 325–342. <https://doi.org/10.1016/j.chb.2015.12.064>
- Raven, J. C. (1958). *Advanced progressive matrices: Set 1*. London, UK: H.K. Lewis.
- Richland, L. E., & McDonough, I. M. (2010). Learning by analogy: Discriminating between potential analogs. *Contemporary Educational Psychology*, 35, 28–43.
- Saldaña, J. (2013). An introduction to codes and coding. In *The coding manual for qualitative researchers* (2nd ed., pp. 1–40). Thousand Oaks, CA: Sage.
- Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, 4, 33–45. <https://doi.org/10.1007/s11409-008-9031-3>
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460–475. <https://doi.org/10.1006/ceps.1994.1033>
- Schraw, G., Dunkle, M. E., Bendixen, L. D., & Roedel, T. D. B. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology*, 87(3), 433–444. <https://doi.org/10.1037/0022-0663.87.3.433>
- Schraw, G., & Gutierrez, A. P. (2015). Metacognitive strategy instruction that highlights the role of monitoring and control processes. In Peña-Ayala, A. (Ed.), *Metacognition: Fundamentals, applications, and trends: A profile of the current state-of-the-art* (pp. 3–16). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-11062-2_1
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7(4), 351–371.
- Schraw, G., & Nietfeld, J. (1998). A further test of the general monitoring skill hypothesis. *Journal of Educational Psychology*, 90(2), 236–248. <https://doi.org/10.1037/0022-0663.90.2.236>
- Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. [https://doi.org/10.1016/0364-0213\(88\)90023-7](https://doi.org/10.1016/0364-0213(88)90023-7)
- Tanner, K. D. (2012). Promoting student metacognition. *CBE—Life Sciences Education*, 11(2), 113–120. <https://doi.org/10.1187/cbe.12-03-0033>
- Thiry, H., Laursen, S. L., & Hunter, A.-B. (2011). What experiences help students become scientists? A comparative study of research and other sources of personal and professional gains for STEM undergraduates. *Journal of Higher Education*, 82(4), 357–388. <https://doi.org/10.1080/00221546.2011.11777209>
- Trey, L., & Khan, S. (2008). How science students can learn about unobservable phenomena using computer-based analogies. *Computers & Education*, 51(2), 519–529. <https://doi.org/10.1016/J.COMPEDU.2007.05.019>
- Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In Wittrock, M. C. (Ed.), *Handbook of research on teaching* (3rd ed., pp. 315–327). New York, NY: Macmillan.
- Weinstein, C. E., & Meyer, D. K. (1991). Cognitive learning strategies and college teaching. *New Directions for Teaching and Learning*, 45, 15–26. <https://doi.org/10.1002/tl.37219914505>
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In Hacker, D. J., & Dunlosky, J. (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Hillsdale, NJ: Erlbaum. <https://doi.org/10.1016/j.chb.2007.09.009>
- Winne, P. H., & Jamieson-Noel, D. (2003). Self-regulating studying by objectives for learning: Students' reports compared to a model. *Contemporary Educational Psychology*, 28(3), 259–276.
- Wittrock, M. C. (1990). Generative processes of comprehension. *Educational Psychologist*, 24(4), 345–376.
- Wittrock, M. C. (1994). Generative Science Teaching. In Fensham P. J., Gunstone R. F., & White R. T. (Eds.), *The content of science: A constructivist approach to its teaching and learning* (pp. 29–38). Bristol, PA: Falmer Press.
- Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology*, 11(4), 307–313. [https://doi.org/10.1016/0361-476X\(86\)90027-5](https://doi.org/10.1016/0361-476X(86)90027-5)
- Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81(3), 329–339. <https://doi.org/10.1037/0022-0663.81.3.329>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(August), 64–70. https://doi.org/10.1207/s15430421tip4102_2