CrossMark

Flipped Classroom Versus Traditional Textbook Instruction: Assessing Accuracy and Mental Effort at Different Levels of Mathematical Complexity

Kristina V. Mattis

Abstract Flipped classrooms are an instructional technology trend mostly incorporated in higher education settings, with growing prominence in high school and middle school (Tucker in Leveraging the power of technology to create student-centered classrooms. Corwin, Thousand Oaks, 2012). Flipped classrooms are meant to effectively combine traditional and online education by utilizing both in and out-of-class time. Despite positively reported implications of the flipped classroom instructional strategy, there is a deep shortage of literature and data that demonstrate advantages for student learning outcomes. The purpose of this preliminary study with directions for future investigations was to examine flipped classroom instruction versus a traditional classroom; specifically, an instructional video versus traditional textbook instruction to assess accuracy and mental effort at three levels of mathematical complexity. College-level nursing students who require mathematical mastery were used as a pilot test group in anticipation that this experience could be translated for larger data sets of variable age groups. Results indicated that accuracy increased and mental effort decreased with flipped instruction. Using Sweller's cognitive load theory and Mayer's cognitive theory of multimedia learning as theoretical frameworks, this study lends insight into designing effective instruction for learning environments that could benefit from a flipped classroom framework.

Keywords Flipped classroom · Mathematics · Accuracy · Mental effort · Nursing students

K. V. Mattis
University of San Francisco, 2350 Turk Boulevard, San Francisco, CA 94118, USA

Present Address:
K. V. Mattis (✉)
212 Miguel Street, San Francisco, CA 94131, USA
e-mail: kristinavmattis@gmail.com

🌀 Springer

1 Introduction

Flipped classroom instruction is a pedagogical strategy primarily used in higher education settings, with growing prominence in high school and middle school (Tucker 2012). Also referred to as the hybrid model (Garnham and Kaleta 2002; Garrison and Kanuka 2004) or blended learning (Morris 2010; Tucker 2012), flipped classrooms convert what would be considered didactical lectures or textbook instruction into how-to instructional tutorials via video or podcast. Through the increase of online instructional videos via YouTube, Curious, Khan Academy, and Vimeo to name a few, or even self-created instruction, educators are transitioning their instructional strategies to include flipped classrooms methodologies. By doing so, instruction is provided outside the traditional four-wall classroom while application of material and assessments are conducted within the classroom in a more kinesthetic and interactive learning approach (Tucker 2012).

Flipped classrooms are meant to effectively combine both traditional and online education by utilizing both in and out-of-class time. Young (2002) reported that most educational environments include a flipped classroom model to offer more efficient learning opportunities especially in situations similar to longer class periods as experienced in a university/college setting or block scheduling in high school. Clark et al. (2006) summarized that efficient instruction, such as flipped classroom instruction, "leads to better learning outcomes with less mental effort," or the amount of cognitive engagement, and that "efficient learning leads to better achievement of instructional goals and/or faster learning" (27). Demetry (2010) indicated that students are far more apt to partaking in collaborative team-based learning in class when lectures are moved outside of class time. Another advantage to flipping classrooms includes facilitating distance learning is to ameliorating recent economic limitations of attending traditional classroom environments, which have led to decreased class sizes and low student–teacher ratios (Berrett 2012).

Even though flipped classrooms provide benefits, instructional strategies are also meant to benefit educational outcomes. Such educational benefits of flipped classrooms are not yet fully understood because there is a deep shortage of literature that demonstrates any possible advantages for student learning and cognitive outcomes. As such, considering the level of complexity of the tasks after receiving the instruction plays a role in the design of the instruction itself.

Complexity is commonly characterized by the number of interactions, or steps, simultaneously taken to solve a problem (Ginns 2005; Mayer 2005a, b). In theory, the more complex a task, the more mental effort is needed to problem solve and the ability to complete the task correctly decreases (Mayer 2005a, b; Paas and van Merrienboër 1993, 1994a). Examining learning materials at different levels of complexity is a way to better assess learning outcomes and cognitive engagement through mental effort. If learners are presented with problem solving tasks that are perceived as challenging, then the task is also considered complex. Tasks can be considered challenging because of lack of experience, prior knowledge, or expertise (Mayer 2009; Sweller 1988). Yet more often than not, time is of the essence and allowing adequate time to assess those reasons can be inefficient. Teachers can act efficiently by accepting that students are in a particular course because of a predetermined level of any of the aforementioned reasons and continue determining the level of complexity of a given task within the course.

The aim of this study was to investigate flipped classroom instruction versus traditional classroom instruction on learning and cognitive outcomes. Learning materials differed in three levels of complexity. Pre- and post-tests were administered to assess gains scores on accuracy for learning outcomes and mental effort for cognitive engagement.

2 Theory

Under the premise of Sweller's (1988) cognitive load theory, Miller's (1956) work on visual and audio working memory channels, and Halford et al. (2005) as an extension of Miller's work, a limited amount of information can be cognitively processed when presented in a single modality such as visual or audio. Single modality presentations force learners to use only one cognitive channel such as listening (e.g. a lecture/podcast) or reading (e.g. a textbook) and, as such, performance can suffer (Costello 2010). Performance can suffer because of cognitive overload during which mental limits of simultaneously processing information are exceeded, and as a result information is not stored accurately or at all (Baddeley and Hitch 1974). Fortunately, Paivio's (1986) dual-coding theory suggests that if material is presented simultaneously in a balanced way across two cognitive channels, a combination of visual and audio, then learners can cognitively process more information and perform better.

A combination of Sweller, Miller, Baddeley and Hitch, and Pavio's theories, Mayer's cognitive theory of multimedia learning encompasses the notion that "people learn more deeply from pictures and spoken words than from pictures and printed words" (2009, p. 200), specifically the modality principle. Under the premise of the modality principle, words are spoken rather than presented as on-screen text. In theory, learners learn better when new information is explained by audio narration than on-screen text. In the very early stages of exploring the modality principle, it was thought that the modality principle worked best when with complex material (Tindall-Ford et al. 1997).

One of the very first positive indications of the principle was presented by a two-part study conducted by Moreno and Mayer (1999) in which participants viewed or listened some combination of on-screen text or audio narration in relation to on-screen animation. Both experiments revealed a modality effect in which participants learned better when words were presented in audio narration rather than on-screen text. The results suggest that the material was complex for the learners, thus presenting a modality effect. In later studies, a modality effect was present when images were complex, the lesson was fast-paced, and the pace was not learner controlled (Clark and Mayer 2011; Tabbers et al. 2004).

Despite previously reported positive implications of the modality principle, Mayer (2009) noted the modality principle was known to work only at high levels of complexity and its effects on low to moderate complexity learning were not well characterized. Even more recently, Schnotz (2011) argued that a modality effect does not work under all conditions. As such, it was found that the modality principle would not likely apply in situations when the presentation exceeds an allotted time, the animation has technical terms or symbols, is not in the learner's native language, or the material is already known to the learner (Clark and Mayer 2011). While there is a difference in presenting stagnant versus moving images, simultaneous presentation of moving visual images and audio in videos is hypothesized to be effective because of the synchronicity. Moreover, and in line with Mayer's spatial contiguity principle, learning will also suffer if the narration and animation are out of sync (Foshay and Silber 2009). While we know how to best design the instructional presentation of the learning materials, we still, however, do not understand if the modality principle presents an effect at levels other than high complexity, thus providing the ultimate premise for this study.

Complexity and problem solving, though not synonymous, appear as a mutual combination. Regardless of Mayer's (2009) suggested limitation, educators have long recognized the necessity to provide students with explicit instruction in complex problem solving

situations (Goldberg and Harvey 1983), and thus, with the recent implementation of flipped classrooms, it is critical to examine the effects of this strategy on learning and cognitive outcomes when the complexity of materials differs. Yet, it is not clear if complexity moderates the modality effect when learning in a flipped setting.

Various researchers have provided descriptive explanations for problem solving. For instance, Chi and Glaser (1985) defined a problem as a situation that covers a large range of difficulty and complexity, requires an end goal, and necessitates finding a way to reach that goal. Then Palumbo (1990) made the connection that the more learners are able to solve problems in realistic and essential situations, the more experience and knowledge they gain. Lastly, Funkhouser and Dennis (1992) regarded problem solving as a process that involves manipulating or operating on previous knowledge in order to find a solution to a problem. Early on, Mayer (1985) also addressed problem solving as a multi-step process that required the problem solver to establish relationships between prior knowledge and the problem at hand with an end goal of successfully implementing a plausible solution.

Based on the extensive meta-analysis conducted by Ginns (2005), studies that incorporated the modality principle were done in mathematical problem solving environments and presented results with strong effect sizes (Atkinson 2005; Ginns 2005; Jeung et al. 1997; Mayer 2009). The National Council of Teachers of Mathematics considers superior problem-solving skills as a way to succeed in mathematics, especially because problem-solving ability is assessed in the classroom and on mandated standardized tests (NAEP 2009; NCTM 2010). An example of such a situation is solving an algebraic equation. Algebraic equations can range in complexity based on the number of elements, or steps, required to successfully solve the problem.

Determining complexity from a cognitive load perspective was first approached by Sweller (1988, 1994), who suggested that complexity could be measured by the amount of steps taken to solve a problem or how many elements interact with one another. Elements are considered single pieces of information. According to Sweller, tasks are low in complexity when elements can be learned in isolation and that a problem can be solved with such isolated elements. Alternatively, if understanding a concept can be done only when simultaneously combining and making connections among several elements, then a task is considered high in complexity. Essentially, complexity levels depend on the number of steps it takes a learner to solve a problem, the conceptual demand in devising a problem, and the number of separate pieces of information needed to be simultaneously processed including the amount of prior knowledge (Mayer 2009; Sweller 1988; Sweller and Cooper 1985). Thus, the level of problem solving is synchronous with the level complexity.

Similar to Sweller's (1988) terminology for complexity, the 2005 National Center for Education Statistics proposed and defined three levels of mathematical complexity: low, moderate, and high. Low complexity relies on recall and recognition of previously learned concepts through which learners could mechanically carry out procedures without an original method or solution. Moderate complexity allows for more flexibility in developing a solution and problems typically have two or more steps where learners are expected to synthesize skill and knowledge from various domains and apply them to the solving process. High complexity places the most demand on learners because the level of engagement in sophisticated abstract reasoning, planning, analysis, judgment, and creative thought.

Like Sweller, Beckmann (2010) also viewed complexity through a cognitive load lens proposing that altering the tasks at hand is a way to change the level of task complexity.

For example, if changing the task at hand alters what needs to be learned, then learning is reflective of essential processing. However, if changing the task at hand does not change what needs to be learned, then learning is reflective of extraneous processing. Beckmann's contemplation directly relates to instructional design regardless of the content domain, but in complexity of that content. One such instructional design is flipping classrooms for math content because math has the potential of examining the instructional design at varying levels of complexity.

Even though there are proposed ways to assess complexity, there is little research on how to quantitatively ascertain complexity levels and how to accurately conclude what is considered more or less complex outside of the cognitive viewpoint (Daniel and Embretson 2010). A potential limitation in assessing complexity is that there is more than one way to solve a mathematical problem, thus impacting the number of steps and conceptual demand to solving problems. Determining cognitive complexity is not often based on empirical or theoretical variables, but "cognitive complexity level is important for measuring both aptitude and achievement" (Daniel and Embretson 2010). Moreover, the lack of empirical research for determining item difficulty within a task prior to the development of such items or before the task is administered leaves many researchers to revert to anticipatory levels of item complexity. For instance, Johar and Ariffin (2001) developed a difficulty index for math problems. The index, however, has rarely been applied to latter studies and thus has little empirical or practical support for its use. Alternatively, the linear logistic test model (Fischer 1995) provides some form of predictability in item difficulty. Similar to the Johar and Ariffin index, however, such models are not routinely applied, especially under the theoretical foundations of this type of study, and thus do not provide sufficient empirical evidence for usability.

Fortunately, complexity levels could be determined by calculating the amount of correct and incorrect answers per problem (Brown 2006). Brown's work consisted of algebraic math problems, problems that vary in complexity in line with Sweller's (1988) definition. For instance, items that were answered correctly 0–33 % of the time indicates a difficult problem, 34–67 % correct indicates a moderately complex problem, and 68–100 % correct indicates an easy problem. Brown's scale also coincides with the sample population in this study; therefore, with the content and population of this study, Brown's scale and Sweller's (1988, 1994) terminology for complexity were adapted for this study.

There is little research examining the modality principle when levels of complexities differ (Ginns 2005; Leahy and Sweller 2011; Mayer 2009; Moreno 2006; Wong et al. 2012). For instance, Jeung et al. (1997) and Mousavi et al. (1995) reported that students receiving video instruction under the modality principle spent less time problem solving and reported higher geometry scores as compared to those receiving visual-only instruction. Yet, complexity of the problems was not considered. Likewise, Costello (2010) reported better general math scores when students received simultaneous visual and audio instruction versus traditional lecture, but there was no indication of performance ability at different levels of complexity when solving such problems. When comparing traditional lecture-based instruction versus flipped classroom, Morris (2010) considered complexity but only on high cognitive load that is typically assumed with high levels of complexity; however, Morris reported that student outcomes suggested benefits using flipped classrooms. Again, though, multiple levels of complexity were not explored.

Given the positive intention of the flipped classroom strategy, there is limited empirical research on examining the impact of learning through such instructional videos on academic achievement and mental effort at different levels of complexity. When applying the modality principle, Mayer's studies suggest that there are positive learning and cognitive

outcomes when task complexity is considered high. As such, theoretically speaking, learning outcomes increase when cognitive load is reduced (Ginns 2005; Mayer 2009), but it is still not known if that holds true when the complexity of the material differs.

In addition to assessing complexity, there were three primary questions in determining the best possible sample for this study: which students would benefit from flipped classroom instruction, which content has different complexity levels, and which students need that specific content to complete course requirements. As such, nursing students require basic algebra knowledge for exit exams and real-world application (Brown 2006). Unfortunately, such students receive little to no algebra instruction because of the time constraints for the core course requirements and the general assumption that older students are already well versed in algebra concepts based on past experience (Cooper and Sweller 1987; Costello 2010; Harrell 1987). Regrettably, undergraduate nursing students experience noticeable deficits in math achievement, specifically algebra concepts such as fractions, decimals, and percentages (Brown 2006; Elliott and Joyce 2004; Gillham and Chu 1995).

Designing math problem-solving items at different levels of complexity is important because it is a way to measure ability and achievement (Daniel and Embretson 2010). Complex algebraic equations are often seen in nursing and require an accurate solution in order to correctly administer medications to patients. This is extremely important because nurses rely on their knowledge and math skills when calculating proper drug administrations for patients. Even though Ginns (2005) and Mayer (2009) reported positive outcomes when using the modality principle in the fields of mathematics, and Costello (2010) and Hodge (2002) examined computer-based instruction within the field of nursing, there is no research specifically investigating the impact of the modality principle on mathematical performance and cognitive impact at different levels of complexity within the field of nursing.

This study aimed at building on prior research by investigating the differences of two instructional strategies—traditional and flipped classroom—on accuracy and mental effort using algebra problems at different levels of complexity: low, moderate, and high on a sample of undergraduate nursing students.

3 Methods and Materials

3.1 Participants

Forty-eight second year university nursing students (22 control; 40 female) from a northern California institution volunteered for this study. Content specialized professors instructed this sample of participants. Small sample sizes of second year nursing students are common with a majority being female (Bull 2009; Costello 2010; Hodge 2002; Melius 2012; Walsh 2008): 63, 70, 40, 51, and 71, respectively. These participants were purposely recruited because nursing students are required to demonstrate understanding of algebra concepts on exit exams and in the workforce. The Institution Review Board for the Protection of Human Subjects accounted for all procedures such that subjects received necessary information to make informed decisions to consent in the participation of this study.

After recruiting participants from a second year nursing program course, participants were quasi-randomly separated into two groups: control and experimental. The control group received traditional textbook instruction on basic algebra concepts involving fractions, decimals, and percentages. The experiment group received the same instruction but

through a video format that included visual and audio modalities. The experiment group followed the concept of a flipped classroom because it received instruction outside of a traditional classroom environment and then returned to the classroom with the knowledge at hand to apply the learned material for assessment purposes.

3.2 Procedures

3.2.1 Pilot of Instrumentation

3.2.1.1 Video Instruction The content of the video was scripted in line with the text-based instruction. With video recordings, however, lighting and sound are critical for clearly conveying information. Lighting includes clear visualization of the information and sound includes clear audio and no extraneous noise. A group of three media experts provided feedback in the development of the video: a professor of multimedia instruction and two classroom teachers with extensive use of video instruction.

3.2.1.2 Math Problems As indicated by Daniel and Embretson (2010), cognitive complexity is not often based on empirical or theoretical variables. As such, to design math problems at different levels of complexity, this study empirically and theoretically applied Mayer's (2009) first definition of determining complexity—the number of steps taken to solve the problem. To solidify different levels of problem complexity, the math problems were first piloted to confirm at least three levels of complexity. A panel of three content experts determined content validity of the math problems: a professor of nursing and two math teachers.

The pilot study included a packet of 18 mathematical problems that fell into one of three categories considered to be areas of deficit as suggested by Brown (2006): decimals (six problems), percentages (five problems), and fractions (seven problems). The pilot took place during the fall semester with 17 participants (n = 15 females) who were enrolled in their first semester of their second year of a nursing program at a northern California university. Second year nursing students were recruited for the pilot study because the sample was most representative of the population for the main study. Even though the study itself would include 15 math problems, the additional three problems in the pilot provided flexibility to re-evaluating problem design if needed. The problems were randomly ordered, rather than by the researcher's anticipated levels of complexity.

There has been debate whether to use numerical equations or word problems when assessing algebraic equation problem-solving ability. Research provides evidence that one of the main concerns when it comes to low problem-solving performance on word problems is the lack of understanding of what the problem is asking (NCTM 2010; Jan and Rodrigues 2012), or the type of language used (Keller 1939; Zakaria and Yusoff 2009; Zakaria 2002). Daniel and Embretson (2010) distinguished that even though equations that are presented in words require more processing steps than numerical-only problems, item difficulty did not increase. Such problems, however, could be associated with increased errors. Alternatively, verbal language from doctors' orders and textual language written in a patient's chart require nurses to understand the specific language that is used. In this study, the word problems were structured in an authentic way that nursing students would see in their work environments. As such, the uses of extraneous words were limited in this study.

A 20-min time limit to complete the packet was in place (Park et al. 2010). In line with cognitive load, McLeod et al. (2003) reported that cognitive processing could be

overloaded if instruction exceeds 20 min and Leahy and Sweller (2011) noted that instruction using the modality principle has more benefit if the lesson has fewer visual images yet longer verbal explanations. Wong et al. (2012) reported that long and complex learning materials should be segmented.

To create a scenario as closely related to the study as possible, the pilot also required the participants to rate their perceived levels of mental effort using the same instrument, but the ratings were not included in determining level of complexity, only accuracy. Accuracy was coded for 0 = incorrect and 1 = correct. Brown (2006) explained that nursing student exit exams require a passing score of 70 %. The logic, thus, in this study was that if 69 % or fewer participants scored correctly on an item, that item was deemed highly complex. If 70–84 % of participants scored correctly, the item was deemed moderately complex. Lastly, if 85 % or more participants scored correctly, the item was deemed low in complexity.

With success, the pilot results of the math problems indicate three levels of complexity: low, moderate, and high. Table 1 provides a breakdown of problems, categories, antici-pated level of complexity and the actual level of complexity.

Table 1 Established complexity levels from pilot to main study

Item	Equation	Category	Anticipated complexity level	Actual complexity level
1	1.67 of 75	Decimals	Low	Moderate
2	0.87 of 94	Decimals	Low	Moderate
3	0.45 of 135	Decimals	Low	Moderate
4	35 % of 80	Percentages	Low	Low
5	9 is what % of 45	Percentages	Low	Low
6	2/5 of 360	Fractions	Low	Low
7	$3.2x + 45 = 6.4$	Decimals	Moderate	Low
8	$34.5 - 6.8x = 45x + 19.81$	Decimals	Moderate	Moderate
9	You have seen 16 of your 20 patients. What percent remains?	Percentages	Moderate	Low
10	1.67 as a fraction	Fractions	Moderate	High
11	12.5 % as a fraction	Fractions	Moderate	High
12	56 % as a fraction	Fractions	Moderate	Low
13	$4.5 - 7.2x = 3.4x - 49.5 + 12.3$	Decimals	High	Low
14	A patient's chart reads that you are to administer 6.95 mls for the first hour and 7.61 mls each hour after. By what percent to you increase the dosage?	Percentages	High	High
15	65 % of 2,000 calories is needed for a 110-pound person. Your patient weighs 76 pounds. What percent of 2,000 calories is needed?	Percentages	High	Moderate
16	You are to administer 1 3/4 l of fluid in 30 min. How many liters in 96 min?	Fractions	High	Moderate
17	If a person weighs in at 72 kg that converts to approximately 159.3 lbs., how many kilograms in 540 lbs.?	Fractions	High	Low
18	You are preparing to administer 1 5/7 mg of antibiotics for every 2 1/3 mg of saline. How many mg of antibiotics will you administer for 18 2/3 of saline?	Fractions	High	High

Two steps were taken to confirm content validity. First, a content expert group consisting of a nursing professor and two math teachers was assembled to ensure that the math problems coincided with each of the three intended categories: fractions, decimals, and percentages. Each individual in the group completed the math problems packet to affirm that each problem matched the intended category and the number of steps needed to solve each problem. Each group member agreed that each problem fit the designated category, and the inter-rater reliability to affirm the number of steps taken to complete each problem was 90 %. While the members were able to complete each problem within 1–2 steps of each other, this reliability level duly notes that people complete math problems differently, some taking fewer steps than others depending on level of expertise. This study did not examine novice versus expertise math levels among the participants.

The second step was to align the complexity level between the pilot study and the main study (Daniel and Embretson 2010). Table 1 charts the anticipated level of complexity as determined by the content validity group. The actual level of complexity was concluded based on the participants' accuracy scores during the pilot study. Because items 13 (decimals) and 17 (fractions) were two complexity levels apart, those two items were omitted from the main study. There were six low, six moderate, and four high complexity problems remaining, and 5 percent, 6 fraction, and 5 decimal problems remaining. To have as balanced a list as possible, the goal was to eliminate either a moderate fraction problem or low fraction problem. This left the option of items 6, 12, and 16. Because items 6 and 12 were both low fractions, it was thought to balance that combination. As such, item 6, a low fraction problem was also omitted from the main study. By doing so, the 15 math problems balanced between the following as charted in Table 2.

3.2.2 Main Study

A pre- and post-test quasi-experimental design was used to measure accuracy and mental effort on three levels of complexity: low, moderate, and high, when solving algebra math problems (Brown 2006; Sweller 1988). Participants were quasi-randomly assigned to either a control or experiment instructional setting. The experiment instruction included an instructional video of worked-out algebraic math examples (Sweller 1988; Paas and van Merrienboër 1994b) with a simultaneous narration explaining the process (Chi and Glaser 1985). For the control group, visual-only instruction was provided designed along the guidelines of a textbook that included text and images in line with Mayer's (2009) spatial contiguity principle where the corresponding text and image are as closely positioned alongside each other, was provided for the control group. The Costello (2010) and Morris (2010) studies lent a basis for the control group as a traditional instructional setting and experiment group as the flipped classroom instruction.

To maintain as much quality control as possible, the pre-test, intervention, and post-test were conducted consecutively without any lapse in time in between (Ginns 2005). Another

Table 2 Balance between categories and levels of complexity among math problems

Categories	Amount	Level of complexity	Amount
Percent	5	Low	5
Fraction	5	Moderate	6
Decimal	5	High	4

reason for the consecutive process was to eliminate as much potential for participants to exempt themselves from the study. A third reason in consecutive presentation of materials was because of the high concern of scheduling conflicts among participants to return to the study setting. While this type of situation is not fully replicated of a typical flipped classroom setting, it was of utmost importance that the already small sample size did not decrease.

Participants from both treatment groups completed a packet of 15 math problems involving algebraic concepts ranging in complexity levels during pre- and post-tests. The types of math problems mimicked the math problems from the pilot study. The pre-test generated a baseline of information to determine gains scores. Learning outcomes were measured by accuracy, and cognitive load was measured by self-perceived mental effort. With a brief introduction and transition times, the procedure lasted 1 h: pre-test 20 min, intervention 13 min, and post-test 20 min. The flipped instruction was recorded at exactly 13 min for the experiment setting; the control setting received the same amount of instructional time with the visual-only format (Wong et al. 2012). While 13 min is considered lengthy, the segmented instruction included a progression from low to high complexity examples of algebra problems, and the high complexity examples requiring longer explanation time (Wong et al. 2012).

3.2.2.1 Control   After the pre-assessment, each participant received visual-only instruction presented on paper that includes text and static images. An example of a math problem that is provided in visual-only format is presented in Fig. 1.

The visual-only instruction was created in accordance with Mayer's (2009) spatial contiguity principle. Previous research on the spatial contiguity principle has indicated that text should be placed as close as possible to the corresponding image in order to avoid a split attention effect on working memory and to eliminate extraneous processing

Decimals

Equation: What is x if 4.5 − 7.2 = 3.4x − 49.5?

1. First, combine like terms

   4.5 − 7.2 = -2.7

2. Now the new equation looks like this

   -2.7 = 3.4x − 49.5

3. Combine like terms

   49.5 + -2.7 = 3.4x

4. Combine like terms again

   Because 2.7 is a negative number, you can either
   49.5 + -2.7
   OR
   49.5 − 2.7.

5. The new equation looks like this

   46.8 = 3.4x

6. Get x to stand by itself by dividing 3.4 from each side

   x = 13.76

Fig. 1 Example of a decimals math problem in visual-only format

(Austin 2009; Mayer 2009). Participants were allowed 13 min to study the instruction. The text for the control was designed to act as the script for the narrated audio in the experimental form of instruction. The intention was to mimic the script from the control as best as possible for the audio. Moreover, the visual images designed for the control were also meant to act as the organizational outline for presenting the information in visual form.

3.2.2.2 Experimental Participants received a 13 min instruction using the modality principle. The instruction was provided as a movie. For this study, instruction was presented at once to the whole group using an LCD projector and audio speakers. To make the instructional video, a document camera was used to record the visual of solving worked out examples of algebraic math problems (Sweller 1988) and narration of a self-explanation (Chi et al. 1994) of the process in real time. The audio mimicked the text version as closely as possible and the worked-out math problems mimicked those in the control group. The following four images demonstrate an example of a decimals problem worked out in chronological order. There is a hand acting as an arrow in images 2 and 3 guiding the viewers to follow the worked out example in sync with the narration.
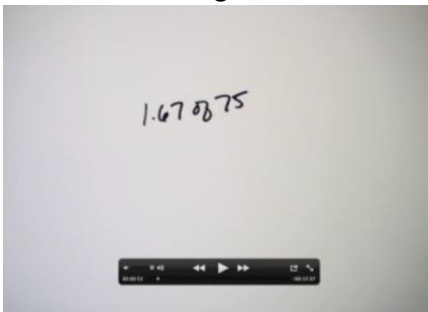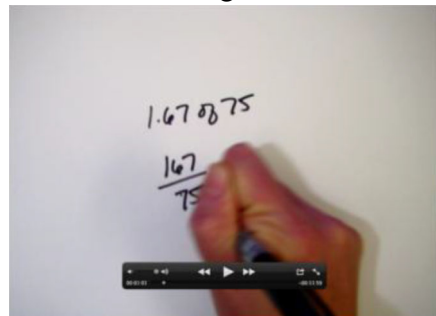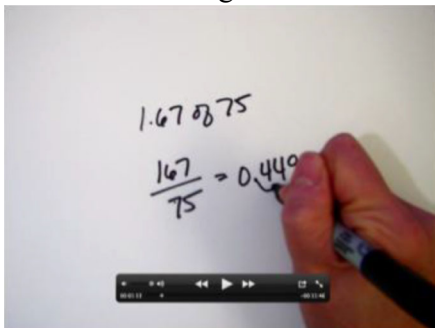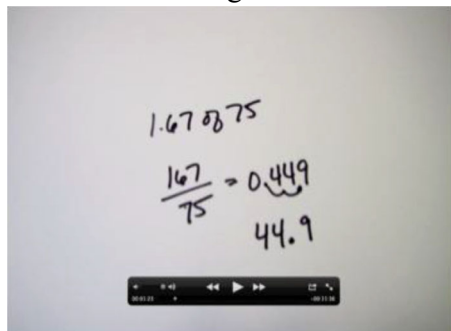
Image 1

Image 2

Image 3

Image 4

After presenting the respective forms of instruction, all participants completed a post-assessment in the allocated 20 min just as for the pre-assessment. After each problem, participants self-reported their perceived level of mental effort.

Accuracy was coded for $0 =$ incorrect and $1 =$ correct. Participants reported self-perceived levels of mental effort on the provided rating scales after each problem during the pre- and post-test through the Perceived Mental Effort Rating Scale (Paas 1992; Paas and van Merrienboër 1993, 1994a). Participants assigned a definitive number to the level of perceived mental effort on a 9-point Likert scale where $1 =$ very, very low perceived mental effort and $9 =$ very, very high perceived mental effort after each item on the pre- and post-assessments. The perceived mental effort assessment provided further information about the cognitive load levels during math problem solving and how to better design instruction to alleviate cognitive load during tasks with varying levels of complexity.

4 Results

The guiding research question for this current study was: What difference does traditional instruction versus flipped classroom instruction have on accuracy and mental effort as the complexity of material changes? Complexity is typically determined by the number of steps required to solve a problem; however, because learners may solve math problems differently albeit correctly, complexity was confirmed by the percent of participants solving correctly (Brown 2006). As indicated by Daniel and Embretson (2010), cognitive complexity is not often based on empirical or theoretical variables. As such, to design math problems at different levels of complexity, this study empirically determined complexity by the level of accuracy as noted in Brown's (2006) study, and by theoretically applying Sweller's (1988) terms of low, moderate, and high to each level. In this study, the math problems were piloted to confirm at least three levels of complexity. Table 3 represents the breakdown of complexity categories based on math problems. Based on the information in Table 3, the breakdown of problems suggests that percentages were of low complexity, decimals were of moderate complexity, and fractions were of high complexity.

ANOVA on repeated measures and Bonferroni corrections were used to conduct analysis on the collected data for accuracy and mental effort on complexity and treatment groups. As in this study, Costello (2010) examined type of instruction on learning outcomes in math with undergraduate nursing students using ANOVA analysis. Because the variables are of focus here, despite the small sample size, Costello's study provides the basis of conducting ANOVA analysis. Furthermore, Costello determined significant main effects using Bonferroni post hoc tests, as the same in this study. Despite similarities with regard to population sample and size, other studies that explored math with undergraduate nursing students noted small sizes, but the purposes and variables differed; thus the analysis procedures from those studies are not accurate for this study (Bull 2009; Hodge 2002; Melius 2012; Walsh 2008).

4.1 Accuracy

Table 4 reports that there was a significant effect of treatment on accuracy at the $p < 0.05$ level for the conditions (F(1, 2) = 0.15, $p = 0.01$).

Specifically, as noted in Table 5, a Bonferroni correction indicates that a statistical significance occurred at the moderate level of complexity ($p = 0.001$, $d = 1.20$) between the control group (M = 0.58, SD = 0.30) and the experiment group (M = 0.85, SD = 0.16).

Table 3 Main study item number, category, and complexity level

Item	Category	Complexity level
1	Decimal	Moderate
2	Decimal	Moderate
3	Decimal	Moderate
4	Percentage	Low
5	Percentage	Low
6	Decimal	Low
7	Decimal	Moderate
8	Percentage	Low
9	Fraction	High
10	Fraction	High
11	Fraction	Low
12	Percentage	High
13	Percentage	Moderate
14	Fraction	Moderate
15	Fraction	High

Table 4 ANOVA on repeated measures accuracy

	SS	df	MS	F	Sig.
Within groups					
Accuracy	1	2	0.50	5.24	0.01*
Accuracy × treatment	0.13	2	0.07	0.69	0.50
Between groups					
Intercept	0.04	1	0.04	0.15	0.70
Treatment	0.08	1	0.08	0.28	0.60

*$p < 0.05$

Table 5 Bonferroni correction for accuracy and treatment

Complexity level	Treatment		T	df	Sig. (2-tailed)	Cohen's d
	Control (n = 22) M (SD)	Experimental (n = 26) M (SD)				
Low	0.73 (0.28)	0.86 (0.18)	1.94	47	0.06	
Moderate	0.58 (0.30)	0.85 (0.16)	3.97	47	0.00*	1.20
High	0.58 (0.37)	0.54 (0.39)	0.36	47	0.72	

*$p < 0.05$

4.2 Mental Effort

As reported in Table 6, there was no significant effect of treatment on mental effort at the $p < 0.05$ level for the conditions (F(1, 2) = 1.89).

However, because the between groups indicated $p = 0.09$, it was of interest to conduct a Bonferroni correction here as well. Table 7 indicates that there was a significant decrease in the use of mental effort ($p = 0.02$, $d = 0.06$) for the participants in the experiment

Table 6 ANOVA on repeated measures—perceived mental effort

	SS	df	MS	F	Sig.
Within groups					
Mental effort	5.37	2	2.68	1.36	0.26
Mental effort × treatment	9.72	2	4.86	2.46	0.09
Between groups					
Intercept	17.46	1	17.46	1.89	0.18
Treatment	23.82	1	23.82	2.58	0.12

Table 7 Bonferroni correction for mental effort and treatment for the experiment group

Complexity level	Treatment		T	df	Sig. (2-tailed)	Cohen's d
	Pre-test (n = 26) M (SD)	Post-test (n = 26) M (SD)				
Low	3.79 (1.53)	3.45 (1.73)	1.12	25	0.28	
Moderate	4.00 (1.70)	3.41 (1.56)	1.78	25	0.09	
High	6.03 (2.33)	4.84 (2.53)	2.55	25	0.02*	0.60

*$p < 0.05$

group on levels of high complexity between pre-test (M = 6.03, SD = 2.33) and post-test (M = 4.84, SD = 2.53).

5 Conclusions and Discussion

The aim of this present study was to examine the impact of flipped classroom strategy use on learning and cognitive outcomes as measured by accuracy and mental effort. Results showed that algebra problems regarded as moderately complex benefit from flipped classroom instruction when concerning learning outcomes. Regarding cognitive outcomes, flipped classroom instruction presented best results when problems were highly complex. While prior literature indicated a connection between lower mental effort and higher academic achievement, this study suggests that even though the highly complex problems were still too difficult to indicate any significant improvement in either treatment group, the benefits of decreased mental effort at the highest complexity resulted in significant score improvement at moderate complexity.

On the pre-test, results indicated that both treatment groups had similar prior knowledge for items of low and high complexity; unexplainably, the experimental group scored higher on items of moderate complexity. Post-test results suggest a significant difference on those same items such that the control group decreased its scores while the experimental group increased its scores, though not significantly.

The results suggest that the experimental group receiving flipped instruction demonstrated higher accuracy on the post-test than the traditional instruction setting, specifically on items of moderate complexity; though none of the results demonstrated a modality effect on such learning outcomes with regard to Mayer's theoretical framework. Thus, results here are in line with Schnotz's (2011) argument that a modality effect does not occur under all conditions. Accuracy differences occurred at levels other than high, unlike

Moreno's (2006) conclusion, perhaps because of gender and learning style preferences. While not a significant change, accuracy decreased for the control group possibly because of Sweller's (1988) split-attention effect.

Moreover, the modality effect may not occur for tasks of low complexity (Paas et al. 2003). Usually, a modality effect occurs when many pieces of information need to be cognitively processed according to Sweller's cognitive load theory; the higher levels of cognitive demand require instruction to free up the limited resources within the visual channel of working memory to decrease potential cognitive overload. On the contrary, this study suggests that significant differences occur at levels of lower cognitive demands.

Though the experimental group initially rated the perceived mental effort much higher than the control group on the pre-assessment, the group largely decreased its ratings on items of high complexity. Coincidentally, the control group increased its ratings, albeit not significantly. It should be noted, however, that both treatment groups scored similarly on items of moderate levels of complexity. There was a significant difference for items of high complexity with a strong effect size. It is unclear as to why there is a significant difference at the high level of complexity given that one would expect similar levels of perceived mental effort on the pre-test prior to any form of instructional treatment.

Pedagogically, it should be noted that math problems categorized as decimals were scored as moderately complex, whereas percentages were low in complexity and fractions were highly complex. When math content involves decimals and fractions, designing instruction using the flipped classroom strategy is suggested. An unintended benefit in this study was that participants were able to authentically experience the video instruction in the environment for which it was meant—outside of class, because the study took place outside of a traditional classroom. This provided further insight into the potential outcomes with instruction that takes place through a distance-learning situation or flipped classroom environments (Garnham and Kaleta 2002; Garrison and Kanuka 2004).

These findings, while focused on improved learning of mathematics for nursing students, could benefit other fields of study such as geometry, chemistry, and biology (Ginns 2005). It should be noted, however, that to examine the learning outcomes and mental effort, instruction materials should include multiple levels of complexity.

6 Limitations and Future Directions

First, it had been imperative to find a sample that required math knowledge and a learning environment that could benefit from a flipped classroom setting. The field of nursing is one such area that encompasses all three aspects of what this study intended. Even though the strength of the results could potentially be weakened by the small sample size, such numbers are common (Bull 2009; Costello 2010; Hodge 2002; Melius 2012; Walsh 2008). Moreover, choosing nonparametric t tests would have been an option for statistical analysis, but because the intention of this study was closely aligned to that of Costello's (2010) study in terms of variables and design of the study, ANOVA analysis was determined. However, this study forms the basis for future research to reconstruct this study throughout multiple locations offering this content specialization.

Second, arguments can be made in choosing a theoretical rationale that may be more practical and more closely aligned to the explicit instructional method of flipping classrooms. However, this study was interested in not only the strategy of flipped classrooms, but also in the learning outcomes and cognitive effects of the instruction. As such, Mayer's theory provides the rationale for the instruction of presenting information via technology

and Sweller's theory provides the rationale for the cognitive component of this study. That said, future studies can incorporate theoretical foundations that are more related on the practitioner scale, and perhaps research that can be closer aligned with statistical findings.

Third, while the flipped classroom instructional strategy is a relatively new concept used in practical learning environments, applying learning and cognitive outcomes to research is not. That said, combining all three concepts together is even newer in approach as indicated by the limited access of prior research. Because of this, locating a leading source of information that provides and publishes such literature coverage is narrow. It is with great hope that as more theoretical researchers and practitioners join forces in exploring twenty-first century approaches in the field of education, more literature will surface.

Fourth, in line with Mayer's theoretical rationale, the modality principle holds strongest for tasks where learners are not able to control the pace of the presentation of instructional materials (Moreno 2006). In the event of distributing the learning material for research purposes of this study, participants in the experiment group received and viewed the learning material one time whereas the control group had the opportunity to re-read material within the allotted time. Typically in non-research settings, students can review instructional videos as much as needed in flipped classroom environments. As this was an unforeseen condition during the study, it begs to question the benefits of self-paced versus instructor-paced learning when flipping classrooms, and the impact on learning and cognitive outcomes.

Fifth, while not an overwhelming limitation, differentiating the results based on gender in this study was not possible because the answers were anonymously submitted. Despite gradual equalization of genders in the field of nursing, the sample in this current study still remained heavily weighted with female learners. It would be of interest to explore the differences in learning and cognitive outcomes in flipped math classrooms with gender as a variable.

Lastly, meaningful learning, according to Mayer (2009) is a result of efficient instruction that produces good transfer and retention on future assessments. In this case, the efficient instruction was in the form of flipped classroom instruction. Post-test results demonstrated a transfer of knowledge gained and retention of such information. Efficient learning is not only reaching the instructional goals as explored in this study, but also faster learning. It would be of interest to measure long-term retention by administering additional post-tests and to measure response times between control and experiment groups in future studies.

References

Atkinson, R. K. (2005). Multimedia learning in mathematics. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 393–408). New York: Cambridge University Press.

Austin, K. A. (2009). Multimedia learning: Cognitive individual differences and display design techniques predict transfer learning with multimedia learning modules. *Computers and Education, 53*, 1339–1354.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Beckmann, J. F. (2010). Taming a beast of burden—On some issues with the conceptualisation and operationalisation of cognitive load. *Learning and Instruction, 20*, 250–264.

Berrett, D. (2012). How "flipping" the classroom can improve the traditional lecture. *The Chronicle of Higher Education*. Feb 19, 2012. https://chronicle.com/article/How-Flipping-the-Classroom/130857/.

Brown, D. L. (2006). Can you do the math? Mathematic competencies of baccalaureate degree nursing students. *Nurse Educator, 31*, 98–100.

Bull, H. (2009). Identifying maths anxiety in student nurses and focusing remedial work. *Journal of Further and Higher Education, 33*(1), 71–81.

Chi, M. T. H., & Glaser, R. (1985). Problem solving ability. In R. J. Sternberg (Ed.), *Human abilities: An information processing approach* (pp. 227–250). San Francisco: Freeman.

Chi, M. T., Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477.

Clark, R. C., & Mayer, R. E. (2011). *E-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (3rd ed.). San Francisco, CA: Wiley.

Clark, R. C., Nguyen, F., & Sweller, J. (2006). *Efficiency in learning: Evidence-based guidelines to manage cognitive load* (p. 27). San Francisco, CA: Wiley.

Cooper, G., & Sweller, J. (1987). The effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*(347), 362.

Costello, M. (2010). *A comparison of three educational strategies for the acquisition of medication calculation skills among baccalaureate nursing students.* Ph.D., Simmons College.

Daniel, R. C., & Embretson, S. E. (2010). Designing cognitive complexity in mathematical problem-solving items. *Applied Psychological Measurement, 34*, 348–364.

Demetry, C. (2010). *Work in progress—An innovation merging "classroom flip" and team-based learning.* Frontiers in Education Conference (FIE), 2010 IEEE. 27–30 Oct 2010, Washington, DC. doi:10.1109/FIE.2010.5673617.

Elliott, M., & Joyce, J. (2004). Mapping drug calculation skills in an undergraduate nursing curriculum. *Nurse Education in Practice, 5*, 225–229.

Fischer, G. H. (1995). *The linear logistic test model: Foundations, recent developments, and applications* (pp. 131–155). New York: Springer.

Foshay, W. R., & Silber, K. H. (2009). *Handbook of improving performance in the workplace, instructional design and training delivery* (1st ed.). San Francisco, CA: Wiley.

Funkhouser, C., & Dennis, J. (1992). The effects of problem-solving software on problem-solving ability. *Journal of Research on Computing in Education, 24*, 338–347.

Garnham, C., & Kaleta, R. (2002). *Introduction to hybrid courses.* Mar, 8. Learning Technology Center, University of Wisconsin-Milwaukee.

Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education, 7*, 95–105.

Gillham, D., & Chu, S. (1995). An analysis of student nurses' medication calculation errors. *Contemporary Nurses, 4*, 61–64.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313–331.

Goldberg, M., Harvey, J. (1983). A nation at risk: The report of the national commission on excellence in education. *Phi Delta Kappan,* 14–18.

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process. *Psychological Science, 16*, 70–76.

Harrell, B. M. (1987). Comparing the effect upon students who are taught nursing math in the classroom with students who do not experience the nursing math course. *RIE*(Jan), 32.

Hodge, J. E. (2002). *The effect of math anxiety, math self-efficacy, and computer assisted instruction on the ability of undergraduate nursing students to calculate drug dosages.* Dissertation, West Virginia University.

Jan, S., & Rodrigues, S. (2012). Students' difficulties in comprehending mathematical word problems in language learning contexts. International Researchers.

Jeung, H. J., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology, 17*, 329–343.

Johar, A. R., & Arrifin, S. R. (2001). *Issues in measurement and evaluation in education.* Bangi: University Kebangsaan Malaysia.

Keller, W. R. (1939). The relative contribution of certain factors to individual differences in algebraic problem solving ability. *The Journal of Experimental Education, 8*, 26–35.

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25*(6), 943–951.

Mayer, R. E. (1985). Implications of cognitive psychology for instruction in mathematical problem solving. In E. A. Silver (Ed.), *Teaching and learning mathematical problem solving: Multiple research perspectives* (pp. 123–138). Hillsdale, NJ: Erlbaum Associates.

Mayer, R. E. (2005a). Multimedia learning: Guiding visuospatial thinking with instructional animation. In P. Shah & A. Miayke (Eds.), *The Cambridge handbook of visuospatial thinking.* New York: Cambridge University Press.

Mayer, R. E. (2005b). Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 169–182). Cambridge: Cambridge University Press.

Mayer, R. E. (Ed.). (2009). *Multimedia learning.* Cambridge: Cambridge University Press.

McLeod, J., Fisher, J., & Hoover, G. (2003). *Key elements of classroom management: Managing time and space, student behavior, and instructional strategies*. Alexandria, VA: ASCD.

Melius, J. (2012). *Math anxiety and mathematics self-efficacy in relation to medication calculation performance in nurses*. Dissertation, University of North Texas.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81–97.

Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted learning, 22*, 149–158.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358–368.

Morris, S. (2010). *A comparison of learning outcomes in a traditional lecture-based versus blended course module using a business simulation with high cognitive load*. Dissertation, University of San Francisco, CA.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reduction cognitive load by mixing auditor and visual presentation modes. *Journal of Educational Psychology, 87*, 319–334.

NAEP. (2009). Nation's Report Card. Retrieved Mar 15, 2010. http://nces.ed.gov/nationsreportcard/mathematics.

NCTM (2010). Overview: Standards for School Mathematics. *Problem Solving.* Retrieved Sept 7, 2010. http://standards.nctm.org/document/chapter3/prob.htm.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skills in statistics: A cognitive load approach. *Journal of Educational Psychology, 84*, 429–434.

Paas, F., & van Merrienboër, J. J. (1993). The efficiency of instructional conditions: An approach to combine perceived mental effort and performance measures. *Human Factors, 35*, 737–743.

Paas, F., & van Merrienboër, J. J. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*, 51–71.

Paas, F., & van Merrienboër, J. J. (1994b). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive load approach. *Journal of Educational Psychology, 86*, 122–133.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1–4.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Palumbo, D. (1990). Programming language/problem-solving research: A review of relevant issues. *Review of Educational Research, 60*, 65–89.

Park, B., Moreno, R., Seufert, T., & Brünken, R. (2010). *Does cognitive load moderate the seductive details effect? A multimedia study*. Paper presented at the American Educational Research Association Conference, Denver, CO.

Schnotz, W. (2011). Colorful bouquets in multimedia research: A closer look at the modality effect. *Zeitschrift fur Padagogische Psychologie, 25*, 269–276.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257–285.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognitive and Instruction, 2*, 59–89.

Tabbers, H. K., Martens, R. L., & van Merrienboër, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology, 74*, 71–81.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*, 257–287.

Tucker, C. R. (2012). *Blended learning in grades 4–12: Leveraging the power of technology to create student-centered classrooms*. Thousand Oaks, CA: Corwin.

Walsh, K. A. (2008). The relationship among mathematics anxiety, beliefs about mathematics, mathematics self-efficacy, and mathematics performance in associate degree nursing students. *Nursing Education Perspectives, 29*(4), 226–229.

Wong, A., Leahy, W., Marcus, N., & Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction, 22*, 449–457.

Young, J. R. (2002). "Hybrid" teaching seeks to end the divide between traditional and online instruction. *Chronicle of Higher Education, 48*(28), A33–A34.

Zakaria, M. J. (2002). *Relationship between learning approach and problem solving on the topic of fraction*. Malaysia: Universiti Kebangsaan Mayalasia.

Zakaria, E., & Yusoff, N. (2009). Attitudes and problem-solving skills in algebra among Malaysian matriculation college students. *European Journal of Social Sciences, 8*, 232–245.