

University of Memphis

University of Memphis Digital Commons

---

Electronic Theses and Dissertations

---

5-5-2023

## Ungrading Writing: Changes in Motivation, Volition, and Perceived Learning

Jennifer B. Wilson

Follow this and additional works at: <https://digitalcommons.memphis.edu/etd>

---

### Recommended Citation

Wilson, Jennifer B., "Ungrading Writing: Changes in Motivation, Volition, and Perceived Learning" (2023). *Electronic Theses and Dissertations*. 3081.

<https://digitalcommons.memphis.edu/etd/3081>

This Dissertation is brought to you for free and open access by University of Memphis Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of University of Memphis Digital Commons. For more information, please contact [khhgerty@memphis.edu](mailto:khhgerty@memphis.edu).

UNGRADING WRITING: CHANGES IN MOTIVATION, VOLITION, AND  
PERCEIVED LEARNING

by

Jennifer B. Wilson

A Dissertation Submitted in Partial Fulfillment of the  
Requirements for the Degree of  
Doctor of Education

Major: Instruction and Curriculum Leadership

The University of Memphis

May 2023

Copyright© Jennifer B. Wilson

All rights reserved.

## Abstract

Motivation and volition are critical precursors to learning, as students learn best when the course material is relevant and personally meaningful. The ARCS-V (i.e., attention, relevance, confidence, satisfaction, and volition) model addresses motivation from an instructional design perspective. Underlying the model is a learner-centered locus of control. Research suggests that grades can be a barrier to learning because they are teacher-centered, allowing instructors to control the reward and punishment cycle. When grades are given, earning high marks often replaces learning as the educational goal. In response to this concern, many undergraduate composition instructors have implemented ungrading, an assessment method in which students are provided with formative feedback without grades. However, an extensive literature search revealed no studies have examined ungrading using the ARCS-V model as a theoretical framework and no studies that examine the ARCS-V model in first-language composition courses. Additionally, few studies have addressed the use of ungrading in online courses. The purpose of this quantitative, survey-based, repeated measures, correlational study was to examine how student motivation, volition, and perceived learning change over time when ungrading is used in online, undergraduate, research and argumentative writing courses (RAW) at a large state university in the Southeastern United States using the ARCS-V model as a theoretical framework. Motivation was measured using the Course Interest Survey (CIS). Volition was measured using the Volition for Learning Scale, and Perceived Learning was measured using the CAP Perceived Learning Scale (CAP Scale). During the spring 2023 semester, 57 students in seven sections of RAW courses participated in an ungraded course. Participants completed the CIS, VFLS, and CAP Scale during week 1 (T<sub>1</sub>), week

4 (T<sub>2</sub>), and week 8 (T<sub>3</sub>) of the course. The results indicated that student motivation remained constant over time and that student volition and perceived learning decreased over time. As these findings contradict much of the existing research, more quantitative research into ungrading is needed to clarify these relationships.

*Keywords:* ungrading, contract grading, labor-based grading, alternative assessment, ARCS model, ARCS-V model, motivation, volition, online learning, distance learning, and composition theory

## Table of Contents

Chapter	Page
List of Tables	vii
List of Figures	ix
List of Abbreviations	x
1. Introduction	1
Problem of Practice Statement	4
Purpose Statement	7
Research Questions	9
Null Hypotheses	10
Significance	12
Definitions	14
2. Review of the Literature	17
Introduction	17
Online Learning	19
Theoretical and Pedagogical Limitations of Ungrading	45
Conclusion	54
3. Methodology	56
Introduction	56
The Investigation Plan	56
Participants	57
Setting	60
Instrumentation	62
Procedures	73
Analysis	75
4. Results	77
Introduction	77
Results	77
Summary	92
5. Discussion	95
Introduction	95
Discussion	96
Implications	108
Limitations and Recommendations	110
Conclusion	113
References	115

Appendix A: Student Assessment Rubric	134
Appendix B: Project 1 - Summary and Response	140
Appendix C: Project 2 - Enter the Conversation	143
Appendix D: Project 3 - Annotated Bibliography	148
Appendix E: Project 4 - Contribute to the Scholarship	153
Appendix F: Course Interest Survey	156
Appendix G: Volition for Learning Scale	159
Appendix H: CAP Perceived Learning Scale	161
Appendix I: IRB Approval	163
Appendix J: Recruitment Announcement	164
Appendix K: Ungrading Handbook	166
Appendix L: Essay Rubric	170
Appendix M: Annotated Bibliography Rubric	173

## List of Tables

Table	Page
1. Summary of Common Ungrading Strategies	26
2. Summary of Instruments	72
3. CIS Total Summary of Descriptive Statistics and Reliability	78
4. CIS Attention Subscale Summary of Descriptive Statistics and Reliability	79
5. CIS Relevance Subscale Summary of Descriptive Statistics and Reliability	81
6. CIS Confidence Subscale Summary of Descriptive Statistics and Reliability	82
7. CIS Satisfaction Subscale Summary of Descriptive Statistics and Reliability	83
8. VFLS Total Summary of Descriptive Statistics and Reliability	84
9. VFLS Volition Planning Summary of Descriptive Statistics and Reliability	85
10. VFLS Volition Control Summary of Descriptive Statistics and Reliability	86
11. CAP Scale Total Summary of Descriptive Statistics and Reliability	87
12. CAP Scale Cognitive Subscale Inter-Item Correlation Matrix	89
13. CAP Scale Cognitive Subscale Summary of Descriptive Statistics and Reliability	89
14. CAP Scale Affective Subscale Summary of Descriptive Statistics and Reliability	90
15. CAP Scale Psychomotor Subscale Summary of Descriptive Statistics and Reliability	91
16. Summary of Normality and Sphericity Assumption Testing	93





## List of Figures

Figure		Page
1.	Psychological Basis of the ARCS-V Motivational Model	28
2.	Sample Age Demographics	58
3.	Sample Racial Demographics	59

## **List of Abbreviations**

Affective Learning Scale (ALS)

American College Testing (ACT)

Attention, Relevance, Confidence, and Satisfaction Model (ARCS)

Attention, Relevance, Confidence, Satisfaction, and Volition Model (ARCS-V)

College Level Examination Program (CLEP)

Confirmatory Factor Analysis (CFA)

Course Interest Survey (CIS)

Desire2Learn (D2L)

English as a Foreign Language (EFL)

Exploratory Factor Analysis (EFA)

Face-to-Face (F2F)

Grade Point Average (GPA)

Institutional Review Board (IRB)

Learning Loss Scale (LLS)

Principal Component Analysis (PCA)

Research and Argumentative Writing (RAW)

Statistical Package for the Social Sciences (SPSS Statistics)

Volition for Learning Scale (VFLS)

## CHAPTER ONE: INTRODUCTION

Theorists and researchers suggest that students learn best when they perceive the instruction as interesting, personally meaningful, and relevant for future success (Angelo, 2017; Franklin, 2017; Goksu & Bolat, 2021; Hobson & Puruhito, 2018). Li and Keller (2018) have argued that motivation and volition are antecedents to learning and are strongly associated with achievement, and most instructors devote considerable time and resources to both motivating students and fostering volition. According to the ARCS-V motivational model, which served as the theoretical framework for the study, both actual and perceived learning increase when motivation and volition are present (Keller, 2010; Li & Keller, 2018). However, promoting student motivation, volition, and perceived learning in online courses is especially challenging, as successful online learning requires affective competencies that many undergraduate students do not possess (Firat et al., 2018; Hobson & Puruhito, 2018; Park & Yun, 2017; Wang et al., 2008). Thus, a need exists for innovative instructional and assessment strategies that foster these attributes.

The ARCS-V model presents an important model to understand intrinsic motivation within online learning contexts. As a systematic design model for improving the “motivational appeal” of instruction, the ARCS-V model addresses the problem of practice from an instructional design perspective (Keller, 1987, p. 2). The foundation of the model is the assumption that people are motivated to engage in activities that satisfy personal needs and where there is a reasonable likelihood of success (Keller, 1987). The original model was comprised of four constructs: attention, relevance, confidence, and satisfaction, which are intended to provide a holistic description of motivation (Keller, 1983). Keller later extended the ARCS model to the ARCS-V model, which includes

volition as an additional construct (Keller, 2008a). Its purpose is to provide a framework for designing motivational instruction that engages students and encourages them to act (Keller, 1979; Keller, 1983; Keller, 1987; Keller, 2008a; Keller, 2008b; Keller, 2010; Keller et al., 2020). Thus, identifying practical strategies for creating effective instruction is the primary objective of the ARCS-V model. As a strategy for increasing motivation (attention, relevance, confidence, and satisfaction), volition, and perceived learning, one could argue that ungrading is well-aligned with this model.

A variety of strategies may be used to support affective learning outcomes, such as motivation. One way to address motivational challenges is through alternative forms of assessment, such as ungrading. As traditionally construed, assessment measures student performance and achievement, often via summative tests. As Lynch and Hennessy (2017) have argued, summative assessment is placed in a “pervasive and privileged” position in higher education, emphasizing “*feedout* to students in the form of grades [with] decreased emphasis granted to engaging students through *feedback*” (p. 1751). They further contended that assessment in higher education is largely focused on “outputs or products” (p. 1754). Conversely, the core feature of ungrading is formative assessment in the absence of numerical scores or letter grades.

Many educational practitioners and researchers have argued that ungrading has a multitude of affective benefits. For example, research findings suggest that it increases intrinsic motivation (Brubaker, 2010; Lindemann & Harbke, 2011), redirects attention from grades to learning (Brubaker, 2010; Chamberlin et al., 2018; Ferguson, 2013; Gorichanaz, 2022; Grau, 1999; Guberman, 2021; McMorran & Ragupathi, 2020; Potts, 2010), improves psychological well-being (Bloodgood et al., 2009; Bullock et al. 2022;

Rohe et al., 2006; Seligman et al. 2021), encourages risk taking (Cowan, 2020; Ferguson, 2013; Gorichanaz, 2022; Guberman, 2021; Mallette & Hawks, 2020; McMorran et al., 2017; McMorran & Ragupathi, 2020), promotes collaboration (Gorichanaz, 2022; McMorran & Ragupathi, 2020; Michaelides & Kirshner, 2005; Potts, 2010; Pulfrey et al., 2011; Roberts & Dorstyn, 2017; White & Fantone, 2010), and enhances autonomy (Brubaker, 2010; Gorichanaz, 2022; Lindemann & Harbke, 2011; Pulfrey et al., 2011). Collectively, the empirical data suggest that students in ungraded courses report a learning orientation (i.e., focused on learning rather than earning high grades) rather than an achievement orientation (i.e., focused on earning high grades rather than learning), increased collaboration and reduced competition, a variety of affective gains, and improved learning as evidenced by improved performance.

The central tenet of ungrading is a focus on assessment *for* learning, or formative assessment, as opposed to assessment *of* learning, or summative assessment (see Black & Wiliam, 1998, Black & Wiliam, 2009; Black et al., 2003, Black et al., 2004). As such, ungrading reflects a paradigm shift regarding the purpose and role of assessment in the learning process. Proponents of ungrading argue that the practice of assigning grades should be abandoned altogether because research suggests that grades divert attention from the learning process and do not tell students how to improve. Additionally, grades can be demotivating for students of all ability levels and can lead to learned helplessness among students who are not high performers (Chamberlin et al., 2018; Gorichanaz, 2022; Guberman, 2021; Mallette & Hawks, 2020; Pulfrey et al., 2011; Tannock, 2017; White & Fantone, 2010). Researchers have noted that ungrading is especially well-suited in certain

disciplines, such as process-based writing instruction, because it allows students to focus on the writing process rather than the final written product (Litterio, 2016).

### **Problem of Practice Statement**

The problem of practice for the study was lack of motivation (attention, relevance, confidence, and satisfaction), volition, and perceived learning among students enrolled in undergraduate, online research and argumentative writing (RAW) courses at a large state university in the Southeastern United States. One problem is that students perceive the course as unnecessary for success in other courses or in their future careers, which diminishes motivation and volition. Because all undergraduate students at the university must complete one semester of RAW, many students enroll in the course merely because it is required. Internal survey data supports this conclusion. During the spring 2019 semester, approximately 1,300 undergraduate students responded to a survey concerning their experiences with general education courses at the university. Results demonstrated that over 76% percent of respondents strongly agreed or somewhat agreed that they would not enroll in general education courses if these courses were not required, and only 30% of respondents strongly agreed or somewhat agreed that general education sources were relevant to their major or useful for their intended careers. Although the survey addressed student opinions of general education courses in general, both anecdotal and student evaluation data from RAW courses supports this conclusion. While the relationship between ungrading and relevance is unclear, relevance is a constituent of the motivation construct in the ARCS-V model, and this relationship was explored in the study.

The problem is not unique to the university of interest, as it is a common concern in all general education courses. As Grau (1999) discussed, mass-enrollment survey courses present unique pedagogical challenges that do not exist in major courses, including lack of enthusiasm and a preoccupation with grades. More recently, Guberman (2021) noted that students perceive only courses in their major fields as contributing to their future goals and that humanities courses were particularly disfavored. Similarly, Spidell and Thelin (2006) remarked that some students appeared “to disdain English classes in particular” (p. 48). These findings are unsurprising, given the marginalization of liberal arts and the trend toward a career-preparation model of higher education. Because students who lack motivation and volition are less likely to learn, instructional and assessment strategies that encourage these characteristics would be beneficial.

Another problem is that some students exit RAW with unsatisfactory proficiency. Departmental assessments have revealed a discrepancy between the current and desired levels of learning, suggesting a need for more effective instructional and assessment strategies. During the spring 2016 semester, the English department at the university of interest assessed RAW capstone essays to identify student competencies after completing the course. A random sample of 104 student essays revealed deficiencies in all learning outcomes with most students scoring at the C-level or below. The grading rubric for the departmental assessment is provided in Appendix A. These results demonstrate a need for instructional methods aimed at increasing learning. Ungrading has been proposed as one potential solution to this problem, as research indicates that it redirects attention from grades to learning by rewarding effort and labor, rather than achievement, allowing students to focus on developing competencies rather than earning high grades (Brubaker,



2010; Chamberlin et al., 2018; Ferguson, 2013; Gorichanaz, 2022; Grau, 1999; Guberman, 2021; McMorran & Ragupathi, 2020; Potts, 2010).

Based on the aforementioned issues, educators have explored ways to support and sustain motivation in online classes. Alternative assessment is one such strategy, as the research literature indicates that grades are a source of dissatisfaction (Carless & Boud, 2018; Denton et al., 2008; Denton & Rowe, 2015; Penn & Wells, 2017; Wood, 2021). Additionally, student evaluations from the university of interest, internal survey data, and verbal conversations with students suggest that some students may perceive general education courses as gatekeepers, rather than foundational components of an undergraduate degree, and that students may perceive poor grades in these courses as punitive. In the literature, students also cite the ostensibly subjective nature of assessment in writing courses as a source of dissatisfaction (Cowan, 2020). As Burger (2017) explained, essays and exams are distinctly different forms of assessment due to the unstructured nature of essay tasks, which prevents “an a priori specification of what constitutes a perfect score,” resulting in a situation where “suspicions of arbitrariness are both more likely to arise and harder to dispel” (p. 305). He further noted that essay assessments can lack transparency. Writing is a complex task that is difficult to evaluate quantitatively, even with the help of a structured rubric (Danielewicz & Elbow, 2009). Although approximately 70% of students at the university earn a passing grade in RAW courses, the pass rate for this course is among lowest in the department.

The research literature, student evaluations, and verbal and written reports from students indicate that lack of motivation (attention, relevance, confidence, and satisfaction), volition, and perceived learning are critical challenges for instructors and

that strategies to increase these characteristics are vital. Although ungrading is not the only potential solution to these problems, administrators within the institution have proposed it as a potential solution, and a substantial minority of instructors now employ ungrading practices. Feedback from student evaluations and classroom discussions about assessment strategies, suggests that ungrading has been successful at increasing motivation (attention, relevance, confidence, and satisfaction), volition, and perceived learning among students in RAW courses. However, these claims have not been empirically assessed, and research is needed to determine their validity.

### **Purpose Statement**

The purpose of the quantitative, survey-based, repeated measures, correlational study was to examine how student motivation (attention, relevance, confidence, and satisfaction), volition, and perceived learning change over time when ungrading is used in online undergraduate RAW courses at the university of interest. A repeated measures design was appropriate because research indicates that although ungrading is associated with increased motivation and volition, these effects may not be stable over time. For example, Pulfrey et al. (2013) distinguished two forms of intrinsic motivation, one focused on task performance and the other based on task persistence, reporting that intrinsic motivation to complete a task does not necessarily lead to continued motivation after the task has been completed. Additionally, student motivation dropped significantly when there were no impending future tasks of the same type, although students in graded condition exhibited a larger decline than students in ungraded condition (Pulfrey et al., 2013). Other researchers have found that student approval of ungrading improves over time. For example, McMorran and Ragupathi (2020) found that 66% of students

approved of ungrading during their first semester, while 72% approved of ungrading during their third semester. Koenka et al. (2019) also noted that motivation is transient and that the effects of feedback on motivation may vary depending on when it is provided relative to task completion. Similarly, Koenka (2020) argued that “it is critical to capture motivation at different assessment stages in order to better understand the influence of performance feedback” (p. 91). Because the relationships between initial and sustained motivation and approval are unclear, more research is needed to explore this relationship in the context of ungrading strategies.

The ARCS-V model served as the theoretical framework for the study. In this model, Keller et al. (2020) defined motivation as “having an objective and developing plans to achieve this objective,” whereas volition involves acting and ultimately achieving the objective (p. 161). Thus, volition follows motivation, as one acts after becoming motivated (Keller et al., 2020). In the study, motivation was measured via scores on the Course Interest Survey (CIS), while volition was measured via scores on the Volition for Learning Scale (VFLS). The third variable was perceived learning. Although course grades often are used to measure achievement, perceived learning, as opposed to actual learning, was used. There are several rationales for this decision. Because all participants were enrolled in ungraded sections of the course, it would have been difficult to assess learning using traditional summative grades. Additionally, research indicates that grades may not be a reliable and valid measurement of learning, as there may be wide variation among the grading scales, assessment types, and criteria used by different instructors, even when teaching the same course (Canfield et al., 2015; Chowdhury, 2018; Gruhlke, 2018; Rovai et al., 2009; Wighting, 2011). Because the study

included two instructors, measuring actual learning based on final course grades, even in an ungraded course, could limit validity due to individual differences between instructors. Finally, course grades often include external factors, such as attendance and participation, which are not indicative of learning (Canfield et al., 2015; Rovai et al., 2009). Thus, a student who has mastered the course content but who exerts limited effort may earn a lower grade than a student who struggles with the content but exerts significant effort (Canfield et al., 2015; Rovai et al., 2009; Wighting, 2011). This factor is especially relevant in ungraded courses, where grades are determined by effort rather than achievement. Given such concerns, Rovai et al. (2009) developed the CAP Perceived Learning Scale (CAP Scale) as a self-report measure of achievement, citing research indicating that students are effective judges of their learning. Thus, for the present study, actual learning was measured as perceived learning via the CAP Scale.

### **Research Questions**

The research questions for this study were:

**Research Question 1.** How does student *motivation* (attention, relevance, confidence, and satisfaction) change over time when ungrading is used in online, undergraduate RAW courses?

**Research Question 1.1.** How does student *attention* change over time when ungrading is used in online, undergraduate RAW courses?

**Research Question 1.2.** How does student *relevance* change over time when ungrading is used in online, undergraduate RAW courses?

**Research Question 1.3.** How does student *confidence* change over time when ungrading is used in online, undergraduate RAW courses?

**Research Question 1.4.** How does student *satisfaction* change over time when ungrading is used in online, undergraduate RAW courses?

**Research Question 2.** How does student *volition* (volition planning and volition control) change over time when ungrading is used in online, undergraduate RAW courses?

**Research Question 2.1.** How does student *volition* change over time when ungrading is used in online, undergraduate RAW courses?

**Research Question 2.2** How does student *volition control* change over time when ungrading is used in online, undergraduate RAW courses?

**Research Question 3.** How does student *perceived learning* (cognitive, affective, psychomotor) change over time when ungrading is used in online, undergraduate RAW courses?

**Research Question 3.1.** How does student *cognitive perceived learning* change over time when ungrading is used in online, undergraduate RAW courses?

**Research Question 3.2.** How does student *affective perceived learning* change over time when ungrading is used in online, undergraduate RAW courses?

**Research Question 3.3.** How does student *psychomotor perceived learning* change over time when ungrading is used in online, undergraduate RAW courses?

### **Null Hypotheses**

The null hypotheses for this study were:

**Null Hypothesis 1.** There is no statistically significant change in student *motivation* (attention, relevance, confidence, and satisfaction) over time when ungrading is used in online, undergraduate RAW courses.

**Null Hypothesis 1.1.** There is no statistically significant change in student *attention* over time when ungrading is used in online, undergraduate RAW courses.

**Null Hypothesis 1.2.** There is no statistically significant change in student *relevance* over time when ungrading is used in online, undergraduate RAW courses.

**Null Hypothesis 1.3.** There is no statistically significant change in student *confidence* over time when ungrading is used in online, undergraduate RAW courses.

**Null Hypothesis 1.4.** There is no statistically significant change in student *satisfaction* over time when ungrading is used in online, undergraduate RAW courses.

**Null Hypothesis 2.** There is no statistically significant change in student *volition* over time when ungrading is used in online, undergraduate RAW courses.

**Null Hypothesis 2.1.** There is no statistically significant change in student *volition planning* over time when ungrading is used in online, undergraduate RAW courses.

**Null Hypothesis 2.2.** There is no statistically significant change in student *volition control* over time when ungrading is used in online, undergraduate RAW courses.

**Null Hypothesis 3.** There is no statistically significant change in student *perceived learning* (cognitive, affective, psychomotor) over time when ungrading is used in online, undergraduate RAW courses.

**Null Hypothesis 3.1.** There is no statistically significant change in student *cognitive perceived learning* over time when ungrading is used in online, undergraduate RAW courses.

**Null Hypothesis 3.2.** There is no statistically significant change in student *affective perceived learning* when ungrading is used in online, undergraduate RAW courses.

**Null Hypothesis 3.3.** There is no statistically significant change in student *psychomotor perceived learning*, as measured by the CAP Perceived Learning Scale (CAP Scale), when ungrading is used in online, undergraduate RAW courses.

### **Significance**

The study was significant in many ways. A critical review of the research literature into ungrading suggests that students consistently demonstrate increased motivation and other affective outcomes (e.g., autonomy, self-efficacy, self-regulation, and psychological wellbeing) when ungrading practices are employed (Bloodgood et al., 2009; Brubaker, 2010; Bullock et al. 2022; Gorichanaz, 2022; Lindemann & Harbke, 2011; Pulfrey et al., 2011; Rohe et al., 2006; Seligman et al. 2021). Additionally, students in ungraded classrooms are more likely to adopt a learning orientation rather than a performance orientation (Brubaker, 2010; Chamberlin et al., 2018; Ferguson, 2013; Gorichanaz, 2022; Grau, 1999; Guberman, 2021; McMorran & Ragupathi, 2020; Potts,

2010), and they engage in more risk-taking (Cowan, 2020; Ferguson, 2013; Gorichanaz, 2022; Guberman, 2021; Mallette & Hawks, 2020; McMorrان et al., 2017; McMorrان & Ragupathi, 2020). Ungrading also is associated with an increased sense of community and collaboration (Gorichanaz, 2022; McMorrان & Ragupathi, 2020; Michaelides & Kirshner, 2005; Potts, 2010; Pulfrey et al., 2011; Roberts & Dorstyn, 2017; White & Fantone, 2010). Finally, ungrading practices are strongly correlated with deep learning (Chamberlin et al., 2018; Gorichanaz, 2022). Considering these benefits, ungrading represents a potentially promising strategy for online college composition courses, as learning activities underlie some aspects of motivation. However, there is a lack of empirical data examining ungrading in online composition courses, and no research studies have addressed ungrading in the context of the ARCS-V motivational model. Thus, studies that blend these elements could provide promising insights. This study attempted to fill this gap.

There were multiple rationales for the study. First, there is a need for more quantitative studies into ungrading, as most of the existing studies have used qualitative methodologies and much of the literature is situated within the personal experiences of individual instructors, often presented in narrative or expository form (e.g., Brilleslyper et al., 2012; Ferguson, 2013; Gomes et al., 2020; Hiller & Hietapelto, 2001; Inoue, 2020; Laflen & Sims, 2021; Mallette & Hawks, 2020; Reardon & Guardado-Menjivar, 2020; Reichert, 2003). As Cowan (2020) noted, composition scholars have only “sporadically” investigated ungrading “using replicable methods” (p. 5). Guberman (2021) characterized this “lack of traditional evidence of effectiveness” as “one of the core challenges facing the expansion of ungrading practices” (p. 89). He called for additional research into



ungrading, arguing that “[u]ltimately, those of us who adopt these methods will be held accountable for providing evidence of student learning and achievement of outcomes” (p. 95). Because empirical data regarding the effectiveness of ungrading is limited, more quantitative research is needed (Chamberlin et al., 2018; Cowan, 2020; Lindemann & Harbke, 2011). Research that addresses the application of the ARCS-V model in composition courses also is needed. In a comprehensive meta-analysis, Goksu and Bolat (2021) found that most studies incorporating the ARCS-V model have taken place in STEM (science, technology, engineering, and math), professional, vocational, and English as a foreign language (EFL) courses. Studies applying the ARCS-V model in EFL courses are plentiful (e.g., Annamalai, 2016; Chang et al., 2016; Chang & Lehman, 2002; Hao, & Lee, 2019; Hung et al., 2013; Jeon, 2021; Kurt & Keçik, 2017; Mirzaei et al., 2022; Proske et al., 2017; Ucar & Kumtepe, 2019). However, a literature search revealed no studies that address the application of the ARCS-V model in first-language composition courses. Thus, research that aligns these two approaches could yield valuable insights with practical applications for educators. The results of the study could be used to evaluate the effectiveness of ungrading in undergraduate RAW writing courses, which has implications for educational practice. Finally, no studies have examined the relationship between ungrading and motivation over time, which is significant given that motivation is not static. The study was an effort to close these gaps.

### **Definitions**

**Attention.** Attention refers to “capturing the interest of learners” and “stimulating the curiosity to learn” (Keller, 2010, p. 45).

**Confidence.** Confidence refers to “positive expectancies for success, experiences of success, and attributions of successes” to ability and effort rather than “luck, chance, or task difficulty” (Keller, 2017, p. 15).

**Extrinsic Motivation.** Extrinsic motivation refers to external motivation where the learner is motivated to achieve a goal because there is a reward or punishment (Deci & Ryan, 1985).

**Intrinsic Motivation.** Intrinsic motivation refers to internal motivation where the learner is motivated to achieve a goal in the absence of an external reward (Deci & Ryan, 1985).

**Motivation.** Motivation refers to “what people desire, what they choose to do, and what they commit to do” (Keller, 2010, p. 3). For the purposes of this study, motivation was operationalized as scores on the CIS.

**Online Learning.** For the purposes of this study, online learning refers to fully-asynchronous distance education.

**Perceived Learning.** Perceived learning refers to student self-reports of learning “within the cognitive, affective, and psychomotor domains” (Rovai et al., 2009, p. 11). For the purposes of this study, actual learning was operationalized as perceived learning as measured by the CAP Scale.

**Relevance.** Relevance refers to “those things which people perceive as instrumental in meeting needs and satisfying personal desires, including the accomplishment of personal goals” (Keller, 2010, p. 48).

**Satisfaction.** Satisfaction “refers to positive feelings about accomplishments and learning experiences” (Keller, 2017, p. 15).

**Traditional Grading.** For the purposes of this study, traditional grading refers to criterion-based or norm-referenced numerical or A-F letter grades.

**Ungrading.** Ungrading is an umbrella term for a variety of alternative assessment strategies that include negotiated grades, grading contracts, labor-based grades, pass-fail grades, and narrative evaluations (Blum, 2020; Cowan, 2020; Gorichanaz, 2022; McMorran & Ragupathi, 2020; Tannock, 2017).

**Volition.** Volition refers to the process of taking definite actions to achieve identified goals (Keller et al., 2020). It involves “the transition from desire, or motivation, to action especially when faced with competing goals” (Keller, 2020, p. 161). For the purposes of this study, volition was operationalized as scores on the VFLS.

## **CHAPTER TWO: REVIEW OF THE LITERATURE**

### **Introduction**

An extensive body of educational research demonstrates that students learn best when they perceive the instruction as interesting, personally meaningful, and relevant for future success (Angelo, 2017; Franklin, 2017; Goksu & Bolat, 2021; Hobson & Puruhito, 2018). These factors are the foundation of virtually all constructivist learning theories (e.g., Deci & Ryan, 1985; Garrison et al., 2000; Merrill, 2002; Zimmerman, 1989; van Merriënboer, 1992), which diverge from traditional didactic educational models in which the instructor is a purveyor of information and a judge of performance rather than a facilitator and a partner in the learning process. Despite the critical role of motivation in the learning process, many students still experience motivational difficulties. This lack of motivation often leads to lack of volition, which inhibits perceived learning (Li & Keller, 2018).

These problems are exacerbated in online courses, with Sapp and Simon (2005) arguing that “online writing courses present both continuing and new concerns about grading” (p. 472). Research shows that students in online courses earn lower grades than students in traditional courses (Bawa, 2016; Sapp & Simon, 2005). Additionally, attrition rates in online courses are quite high. For example, Sapp & Simon (2005) found that 30% of students did not complete their online composition courses. Kearsley (2002) asserted that online students must “show extraordinary self-discipline and initiative,” abilities that many students in first-year composition courses do not possess (p. 476). Motivation is a key precursor for these characteristics. As Bawa (2016) concluded in a literature review on retention in online courses, motivation often is the deciding factor for retention among

online students. Thus, assessment methods in these courses should be designed to encourage the affective traits necessary for student success.

In the research literature, ungrading has been proposed as a potential strategy for increasing motivation (attention, relevance, confidence, and satisfaction), volition, and perceived learning. The fundamental assumption of ungrading is that extensive formative feedback should be substituted for summative grades. The importance of formative assessment is well-established, as studies show that students rely on timely, effective, actionable feedback that analyzes their strengths and weaknesses (Nicol & McFarlane-Dick, 2006; Pan & Shao, 2020; Penn & Wells, 2017; Wang & Zhang, 2020). However, students consistently identify feedback and assessment as sources of dissatisfaction (Carless & Boud, 2018; Denton et al., 2008; Denton & Rowe, 2015; Penn & Wells, 2017; Wood, 2021). Thus, strategies aimed at improving the quality of feedback are needed. Ungrading is one alternative strategy for achieving these aims and addressing these issues, especially in the context of online learning.

While the ARCS-V model has been extensively studied in other domains, no studies have used ungrading as an instructional intervention in the context of the ARCS-V model. Similarly, although the model has frequently appeared in studies conducted in EFL courses (e.g., Annamalai, 2016; Chang et al., 2016; Chang & Lehman, 2002; Hao, & Lee, 2019; Hung et al., 2013; Jeon, 2021; Kurt & Keçik, 2017; Mirzaei et al., 2022; Proske et al., 2017; Ucar & Kumtepe, 2019), the literature search returned no studies in which the ARCS-V model was used in first-language composition courses. This underrepresentation suggests a need for more studies that incorporate the ARCS-V model in these courses. These gaps in the research formed the foundation for the study.

This review is divided into three sections. First, the theoretical context for the review is outlined, especially in the context of online learning. This section is followed by a review of the literature into the merits of ungrading, including a description of the pedagogical strategies used in ungrading and the benefits of ungrading as presented in the research literature. Finally, the pedagogical and theoretical limitations of ungrading are examined.

### **Online Learning**

The ARCS-V model frequently is used in research involving online learning, and research suggests that online learning is more than a mere addition to the traditional classroom experience; it is a distinct entity, requiring new teaching strategies (pedagogy and design) to accommodate its unique affordances (Franklin, 2017; Garrison & Anderson, 2003; Neal, 2011; Stella & Corry, 2013). As Franklin (2017) argued, “Online teaching and learning present a variety of issues and questions that research from traditional classrooms cannot fully address (p. 51). Many of the benefits of online learning are self-evident. For example, online learning has democratized education, providing access to those for whom the traditional classroom experience is unfeasible. It also allows learners to communicate and share resources with a wider audience and fosters more thoughtful participation through online discussion boards.

Despite the promise of online learning, however, it has heralded new problems that reflect the limitations, especially in asynchronous communication, which can impact affective learning outcomes (Anderson et al., 2001; Berge & Collins, 1995; Garrison et al., 2000; Sapp & Simon, 2005). Garrison et al. (2000) characterized computer-mediated communication as a “lean medium” that filters out “non-verbal or paralinguistic cues

such as facial expression and tone of voice” in fully-asynchronous courses (p. 90).

Providing effective formative feedback can be difficult in all educational settings.

However, it is particularly challenging in online courses due to the inherent limitations of asynchronous communication (Anderson et al. 2001; Berge & Collins, 1995; Garrison et al., 2000). Thus, online learning presents new pedagogical challenges that require innovative solutions, and instructors must continuously seek out teaching (pedagogy and design) and assessment strategies to mitigate these barriers and support the affective component of learning.

### **Role of Formative Assessment for Online Learning in Composition Courses**

Formative assessment is a critical component of composition pedagogy, which endorses an integrative approach to writing. Most online composition classrooms are structured as writing communities in which students work collaboratively to complete multimodal, problem-based assignments (Borg, 2003; Horstmanshof & Brownie, 2013; Lam, 2016; Lam, 2018; Swales, 2016). This paradigm endorses a process-oriented approach to writing in which the final product is less important than the steps taken to produce the product (Murray, 2003). Thus, writing tasks are presented in sequence, with subsequent tasks building on previous work. Because feedback is a critical component of constructivist theories, instructors in these online courses are tasked with providing effective and efficient feedback that will help students to meet learning objectives. Given this process approach, formative feedback is especially important, and frequent, incremental feedback typically is built into the teaching and learning process (Murray, 2003; Riddell, 2015).

Despite its importance, providing effective formative feedback is a continual challenge in composition courses (Connors & Lunsford, 1993; Lynn, 2004). One problem is the assessment model in these types of online courses, which poses concerns because patently correct or incorrect answers do not exist within this domain. Thus, grades in these courses are somewhat subjective, and the grading criteria can be ambiguous for students (Cowan, 2020; Koenka et al. 2019). Balancing affective issues also can be incredibly challenging in writing courses. Noting that writing is more personal than work in other disciplines, Sapp and Simon (2005) have asserted that “the tenor of instructor critique requires much more attention from writing teachers than from teachers of other subject matter” (p. 479). Indeed, this is especially true in the online environment where emotional cues are limited (Neal, 2011).

The amount and type of feedback in these online courses are important as well. Extensive feedback can overwhelm and discourage students, while insufficient feedback is unlikely to produce meaningful learning (Butler, 1987; Swaffield, 2011). Thus, instructors are tasked with providing feedback that is constructive, helpful, and detailed without overwhelming students. This task can be difficult because composition instructors must address an array of complex issues, including critical thinking, organization, development, style, research, grammar, and digital literacy (Adler-Kassner & O’Neill, 2010; Huot, 2002; Weigle, 2009). As Cowan (2020) has argued, “A single grade hardly sums up the complicated mixture of strengths and weaknesses present in students’ work, and if an instructor’s goal is to teach invention, feedback, revision, and other process-oriented skills, grading one final product misses the point entirely” (p. 1). Based on this research, one might argue that alternative assessment strategies are needed



to mitigate the negative consequences of summative grades in online composition courses.

### **Ungrading as Alternative Formative Assessment in Online Learning**

The practice of assigning grades has been controversial for decades, with critics arguing that grades emphasize performance over learning (Brubaker, 2010; Chamberlin et al., 2018; Ferguson, 2013; Gorichanaz, 2022; Grau, 1999; Guberman, 2021; McMorran & Ragupathi, 2020; Potts, 2010). In his seminal article on feedback, Ramaprasad (1983) argued that feedback is useful only when it identifies a learning gap and provides information that will allow learners to close this gap. In another extensively cited study, Butler (1987) found that students often ignore formative feedback when it is accompanied by grades, a finding that has been reinforced by more recent studies (e.g., Koenka, 2020; Koenka et al., 2019; Pulfrey et al., 2011). Nearly forty years after these articles, however, grades remain relatively ubiquitous across learning institutions.

Given the criticisms of traditional grades, alternative assessment strategies are needed. Ungrading is not a recent development. It appeared prominently in scholarly literature during the 1970s (e.g., Burton et al., 1977; Martin & Harrison, 1972; Polczynski & Shirland, 1977; Stasz, 1976; Stelzner 1975; Wolvin & Wolvin, 1975; Yarber, 1974). Modern research on ungrading in composition courses extends from the work of Peter Elbow (1968, 1983, 1993, 1997a, 1997b), Asao Inoue (2004, 2012a, 2012b, 2014, 2019, 2020), Alfie Kohn (1994, 1999, 2002, 2011), and Ira Shor (1996, 2009), who are strong opponents of traditional grading. Its resurgence parallels the gradual transition toward constructivism, critical theory, and postmodernism in education

and online learning mediums. It also coincides with movements to ban high stakes testing in schools (Tannock, 2017).

### ***Types of Ungrading Formative Assessment Strategies***

Formative assessment is the keystone of ungrading pedagogy. Seminal research into formative assessment conducted by Butler (1987) suggests that formative feedback alone is more effective than summative feedback or a combination of formative and summative feedback. This research also indicated that low achievers exhibit low interest when grades are given. Most work on formative assessment stems from the work of Black and Wiliam (1998), who characterize assessment as a moment of contingency, acknowledging the fluid and transitory nature of knowledge and skills development. Their intentionally broad definition encompasses “all those activities undertaken by teachers, and/or by their students that provide information that can be used as feedback to modify the teaching and learning activities in which they are engaged” (Black & Wiliam, 1998, p. 8). They also stressed the importance of providing *learning* goals rather than *performance* goals, noting that the primary assumption of formative assessment is that all students can and will achieve (Black et al., 2004). As Soles (2001) related, students learn best when formative feedback identifies “those aspects of their work that deserve praise and criticize[s] weaknesses constructively, in a voice that suggests that the student’s work can and will improve with some extra effort” (p. 123). Although “learning completely delinked from letter and number grades” is the ideal form of ungrading, it is uncommon in practice, as most universities require instructors to assign final grades (McMorran & Ragupathi, 2020, p. 926).

Ungrading does not imply that there are no assessments, merely that assessment is divorced from abstract notions of achievement and performance (Laflen & Sims, 2021). Instructors still address the quality of student work in their feedback, and students in ungraded courses are expected to complete assignments and to devote time and effort to their studies (Ferguson, 2013). Ungrading pedagogy, thus, adopts a mastery learning approach (see Bloom, 1971). When the quality of student writing falls below minimum standards of competence, students are asked to revise and resubmit those assignments until they demonstrate competency. Those who fail to meaningfully participate in the revision process must repeat the course. The primary divergence from traditional assessment is that effort and progress displace achievement as the educational goal. As Laflen and Sims (2021) explained, the distinction is that ungrading involves “meeting collaboratively determined standards for labor rather than meeting some pre-determined, subjective standard of proficiency” (p. 120).

Ungrading has become an umbrella term for a variety of alternative assessment strategies that include the following: narrative evaluations, pass-fail grades, negotiated grades, grading contracts, and labor-based grades. Instructors using *narrative evaluations* assign no grades whatsoever (Guberman, 2021; Marshall et al., 2005; Tannock, 2017). This form of ungrading is preferable, as it completely separates grades from the learning process. However, narrative evaluations are uncommon in higher education because most colleges and universities require instructors to assign a final grade at the end of the term. Instructors may remedy this problem by adopting modified versions of ungrading.

The most common compromise is *pass-fail grading* in which grades are not awarded according to the traditional letter grade system. This form of ungrading most

resembles traditional grading because it merely collapses the category gradations from five (i.e., A to F) to two (i.e., pass or fail). There are two forms of pass-fail grading. First, individual assignments may be traditionally-graded, but students are assigned either a passing or a failing final grade for each course. This is known as hybrid pass-fail grading because numerical grades are used to determine the pass or fail designation (McMorran & Ragupathi, 2020). In the second form of pass-fail grading, individual assignments are evaluated as pass or fail, as is the final grade for the course (McMorran & Ragupathi, 2020).

Another form of ungrading includes strategies whereby instructors and students are partners in the grading process. *Negotiated grading*, also referred to as participatory assessment, refers to a strategy in which the instructor and the student meet at the end of the semester to review the student's progress in the course (Brilleslyper et al., 2012; Tannock, 2017). Both the instructor and the student propose a grade that reflects the student's performance in the class (Brilleslyper et al., 2012; Gorichanaz, 2022; Tannock, 2017) If the instructor and student disagree about the proposed grade, then they negotiate a grade that is mutually agreeable (Brilleslyper et al., 2012, Guberman, 2021). In an effort to democratize assessment and support affective learning outcomes, instructors using this method often defer to students when their grade differs from that of the instructor (Gorichanaz, 2022).

Like negotiated grades, *grading contracts* are designed to involve students in the assessment process. In this form of ungrading, the instructor creates a contract that sets forth the minimum requirements for each letter grade (Cowan, 2020; Elkins, 2016; Gomes et al., 2020; Mallette & Hawks, 2020; Reardon & Guardado-Menjivar, 2020). At

the beginning of the semester, students sign a contract for a particular grade and agree to complete the assignments required to earn that grade (Cowan, 2020; Gomes et al., 2020; Inoue, 2020; Laflen & Sims, 2021; Mallette & Hawks, 2020; Reardon & Guardado-Menjivar, 2020). For example, the contract might state that students must complete all major projects to earn a passing grade in the course, while they must complete all process assignments and discussion posts to earn a higher grade. Typically, students may renegotiate the contract as the semester progresses (Cowan, 2020; Hiller & Hietapelto, 2001; Mallette & Hawks, 2020).

Finally, in *labor-based grading*, instructors assign final grades based on the number of assignments completed (Gomes et al., 2020; Laflen & Sims, 2021). Apart from the requirement that students must revise assignments until they reflect a minimum level of competence, grades are not awarded based on quality, and there are no letter or numerical grades on any assignments (Laflen & Sims, 2021). The key distinction is that final grades are awarded based on labor, rather than achievement, in labor-based grading. Students receive an evaluation of *accept* or *revise* for individual assignments. Merely completing all assignments at the *accept* level earns students a letter grade of A in the course. Many instructors combine labor-based grading with grading contracts to create a labor-based grading contract (Cowan, 2020; Gomes et al., 2020; Inoue, 2020; Laflen & Sims, 2021; Reardon & Guardado-Menjivar, 2020).

**Table 1**

*Summary of Common Ungrading Strategies*

Ungrading Strategy	Description
Narrative evaluations	Instructors assign no grades whatsoever. All feedback (both summative and formative) is narrative. Ideal form of ungrading but uncommon in practice.

Pass-fail grades	Instructors assign a grade of <i>pass</i> or <i>fail</i> with no numerical or multi-interval letter grades.
Negotiated grades	The instructor and student meet at the end of the course to collaboratively negotiate a mutually agreeable grade. Instructors typically defer to the student when there is a discrepancy.
Grading contracts	Grades are assigned based on the terms of a contract that sets forth the requirement for each grade interval. Students select a desired grade and agree (in writing) to complete the requirements for that grade. Students typically may renegotiate for a different desired grade once the course is underway.
Labor-based grades	Grades are assigned based on assignment completion. Individual assignments are designated as <i>accept</i> or <i>revise</i> . Students revise assignments until they meet minimum competence. Full credit is assigned for meeting minimum competence (i.e., a designation of <i>accept</i> ).

---

### **Ungrading and the ARCS-V Model for Online Learning**

One way to further understand ungrading is to examine how it may benefit affective learning outcomes, such as motivation. Motivation is multifaceted and can be broadly classified as either intrinsic or extrinsic. Intrinsic motivation refers to internal motivation where the learner is motivated to achieve a goal in the absence of an external reward (see Deci & Ryan, 1985). Thus, the learner perceives the goal or activity as rewarding in itself. It is important to note that subject matter interest and enjoyment are not necessary prerequisites for intrinsic motivation. As ChanMim et al. (2017) argued, students may be intrinsically motivated by other factors. For example, medical students may not find studying human anatomy to be interesting or enjoyable, but they could still be intrinsically motivated to do so because they recognize that this knowledge is relevant for their future careers (ChanMim et al., 2017). Conversely, extrinsic motivation refers to external motivation where the learner is motivated to achieve a goal because there is a reward or punishment (see Deci & Ryan, 1985). One goal of the ARCS-V model is to

promote intrinsic motivation from an instructional design perspective. Hence, the ARCS-V constructs help guide the design of learning experiences and materials for educators who wish to support intrinsic motivation in their online courses.

Especially in online learning contexts, motivation may be operationalized through the ARCS-V model, which is intended to meet the motivational (attention, relevance, confidence, and satisfaction), volitional, and perceived learning needs of a diverse population of students (Keller, 2008a; Keller, 2008b; Keller, 2016; Li & Keller, 2018). The model includes four motivational constructs: attention, relevance, confidence, and satisfaction. The fifth construct, volition, is distinct from the motivational constructs because it represents the transition from mental processes to action (Keller, 2010). The four motivational constructs involve identifying objectives and developing plans to achieve them. Figure 1 depicts the theoretical constituents of the ARCS-V model.



## **Figure 1**

### *Psychological Basis of the ARCS-V Motivational Model*

*Note.* From “Effects of the ARCS-V-Based Motivational Strategies on Online Learners’ Academic Performance, Motivation, Volition, and Course Interest,” by H. Ucar and A. T. Kumtepe, 2019, *Journal of Computer Assisted Learning*, 36, p. 336.

(<https://doi.org/10.1111/jcal.12404>). 2019 John Wiley & Sons Ltd.

### ***Attention***

Attention refers to “capturing the interest of learners” and “stimulating the curiosity to learn” (Keller, 2010, p. 45). Keller (2017) explained that attention incorporates “curiosity and arousal, interest, boredom, and other related areas such as sensation seeking” (p. 14). He characterized attention as a necessary precursor to learning, arguing that no learning can take place without engaging the learner’s attention (Keller, 2010). Keller (2010) also has indicated that curiosity is a critical factor in sustaining attention. He referred to the necessary type of curiosity as epistemic curiosity, which occurs when one fosters a desire for answers or a desire to learn new subjects (Keller, 2010). Attention is an important component of intrinsic motivation. Indeed, in a study of 62 graduate students at a public university in the Southeast, Novak (2014) found that attention is the greatest predictor of student satisfaction and that course interest is closely aligned with satisfaction. Keller and Suzuki (2004) further explained that “attention and curiosity are necessary, but not sufficient, conditions for motivation” (p. 231). Instructors also must sustain attention, which requires “incongruity or conflict” and “variability” (Keller & Suzuki, 2004, p. 231). Keller (2010) identified three strategies for fostering attention: perceptual arousal, inquiry arousal, and variability. Perceptual arousal



involves capturing interest and arousing curiosity (Keller, 2010). Inquiry arousal refers to increasing interest once it has been obtained (Keller, 2010). Finally, variability is aimed at maintaining attention (Keller, 2010). All three forms of attention are necessary components of motivational instruction.

Few researchers have targeted the attention construct in isolation, as most studies address all four motivation constructs. For example, in a video content analysis study of 12 experienced EFL teachers in Korea, Jeon (2020) found that teachers used significantly more attention strategies ( $n = 354$ ) compared to strategies designed to foster relevance ( $n = 131$ ), confidence ( $n = 144$ ), or satisfaction ( $n = 149$ ). Jeon (2020) explains that the most commonly used attention strategy was variability, as the experienced instructors devoted constant effort to “make their lessons more interesting by changing the instructional style and activity types” (p. 269). In terms of learning design and strategies, the key finding of this research is that instructors must gain learners’ attention by introducing novel or contradictory concepts and sustain attention by varying instructional strategies (Keller & Suzuki, 2004). Thus, attention must be both aroused and maintained within design for it to support learning outcomes.

### ***Ungrading and Attention***

As defined in the ARCS-V model, attention includes curiosity and course interest. Although grades can increase short term motivation, researchers have argued that grades do not promote attention, course interest, or long-term retention (Chamberlin et al., 2018). These mental states are precursors to the attention construct of the ARCS-V model. Instead of active attention, students in graded courses tend to adopt a superficial approach in which information is discarded once the associated summative assessment

task has been completed (Chamberlin, 2018). By contrast, ungrading research suggests that this type of alternative assessment fosters attention and course interest. For example, in an interpretative phenomenological interview study conducted with nine undergraduate information technology students at Drexel University, Gorichanaz (2022) found that negotiated grades increased curiosity, an element of attention, as well as metacognition. He further argued that when students are learning oriented due to ungrading, as opposed to grade oriented, they are more interested in completing course assignments and perceived that they had more freedom to choose personally meaningful topics and to experiment with unfamiliar composition strategies when completing these assignments (Gorichanaz, 2022). Similarly, McMorran et al. (2017) discussed that “developing intrinsic reasons for learning, such as for the joy of learning or the development of skills,” is impeded when grades are assigned (p. 362). Intrinsic motivation leads to course interest. Therefore, course interest is an important component of the attention construct.

### ***Relevance***

Keller (2010) defines relevance as “people’s feelings or perceptions of attraction toward desired outcomes, ideas, or other people based upon their own goals, motives, and values” (p. 98). To sustain motivation, learners must perceive the instruction as relevant. Keller (2008b) described relevance as a perception that the “knowledge to be learned is perceived to be meaningfully related to a learner’s goals (p. 177). He argued that students often view relevance from a pragmatic perspective and that they “most often want to be told or shown how [the course material] will be useful to them in their jobs or have a practical application in some other part of their lives,” while acknowledging that this is not always possible (Keller, 2010, p. 98). Like motivation, goals may be extrinsic or

intrinsic (Keller & Suzuki, 2004). As Hobson and Puruhito (2018) have explained, “When students fail to see how they can utilize the information they are learning for their future beyond simply receiving a passing grade that allows them to move one step closer to graduation, learning and performance suffer” (p. 137). They further argued that instructors should embed relevance-building opportunities into instructional design so that students can connect the course material to their future careers (Hobson & Puruhito, 2018).

Although extrinsic goals, such as the desire to pass a course or earn a degree, can stimulate relevance, sustaining relevance requires goals that are personally meaningful (Keller & Suzuki, 2004). The primary method for creating this type of relevance is by clearly relating the instruction to learners’ future objectives, either personal or professional (Keller & Suzuki, 2004). However, tying instruction to solely career goals may be initially necessary, but insufficient, for achieving long-term relevance (Franklin, 2017). It also involves conveying “the innate value of the learning experience” (Franklin, 2017, p. 48).

Keller (2010) identified three strategies for increasing relevance: goal orientation, motive matching, and familiarity. Goal matching involves providing verbal statements and examples of the utility of the course material (Keller, 2010). Motive matching involves providing achievement opportunities, opportunities for collaboration, leadership responsibilities, and positive role models (Keller, 2010). Finally, familiarity involves providing examples and analogies that relate to the learner’s background (Keller, 2010).

### *Ungrading and Relevance*

The relationship between ungrading and relevance is unclear, as it is mostly absent from the literature. However, some researchers have specifically examined the role of relevance in the ARCS-V model. For example, Means et al. (1997) argued that relevance is the most important precursor to learning and that relevance should be the primary focus of motivational research, citing evidence that students perform better when they perceive the course content to be personally meaningful. In their experimental study of 100 undergraduate students enrolled in statistics and human physiology courses, they concluded that when the course material is not intrinsically relevant to students, instructors must include relevance-enhancing strategies if students are to achieve learning outcomes, arguing that such strategies are more effective than other embedded strategies aimed at increasing motivation (Means et al., 1997). By adopting a learning orientation rather than an achievement orientation, ungrading encourages students to move beyond external rewards towards greater self-regulation and the development of personally meaningful learning objectives.

Chang and Lehman (2002) reported similar conclusions with 313 undergraduate EFL students at a university in China, arguing that educators heavily rely on external motivation and that students are motivated to please teachers and parents rather than developing “an internal thirst for knowledge and experience” that relates to their experience (2002). Similarly, Chamberlin et al. (2018) found that students’ interest in the subject wanes following the assessment task, especially with subjects and tasks that are not intrinsically interesting to them (Chamberlin et al., 2018). In addition to encouraging deeper learning, ungrading frees students to take courses that are relevant to them rather

than courses in which they excel, which McMorran et al. (2017) have characterized as a “rare opportunity to finally break free from only pursuing subjects they know they excel at, and instead expand one’s learning horizons” (p. 370). However, none of these studies directly examined the relationship between ungrading and relevance. In part, the purpose of the study was to examine this relationship, which is especially important in undergraduate general education courses, as most students enroll in these courses merely because they are required.

### ***Confidence***

Confidence refers to “positive expectancies for success, experiences of success, and attributions of successes” to ability and effort rather than “luck, chance, or task difficulty” (Keller, 2017, p. 15). Keller and Suzuki (1988) delineated three dimensions of confidence: perceived competence, perceived control, and expectancy for success. Perceived competence refers to learners’ beliefs that they *can* succeed (Huett et al., 2008; Keller & Suzuki, 1988). Students are unlikely to engage in activities when there is little likelihood of success (Huett et al., 2008; Keller & Suzuki, 1988). Perceived control relates to learners’ beliefs that their success depends upon their own abilities and efforts (Huett et al., 2008; Keller & Suzuki, 1988). When learners perceive their success to be the result of external factors, confidence diminishes (Keller & Suzuki, 2004). Thus, learners must attribute their successes to their own actions. Finally, expectancy for success refers to learners’ beliefs that they *will* succeed (Huett et al., 2008; Keller & Suzuki, 1988).

Keller (2010) defines three strategies for increasing confidence: learning opportunities, success opportunities, and personal control. Providing learning

opportunities involves establishing trust and positive expectations for success (Keller, 2010). Providing success opportunities involves incorporating assignments and experiences that lead to success, thereby increasing learners' success expectations on future assignments (Keller, 2010). Finally, personal control involves conveying to students that their success is dependent on their own efforts rather than external factors (Keller, 2010). Keller (2016) argued that confidence can be enhanced by “creating success expectations, providing opportunities for success, and instilling personal responsibility for success” (p. 5). As Milman and Wessmiller (2020) have contended, formative feedback that emphasizes strengths and targets weaknesses with suggestions for improvement is critical for increasing confidence.

In the literature, confidence has been addressed in online courses through various means. A literature search suggests only two recent studies have targeted the confidence construct of the ARCS-V model independently. In an experimental study of 81 undergraduate students at a state university in Texas, Huett (2006) examined the effects of confidence-building tactics – such as motivational emails, simple feedback on assessment tasks, and self-paced assessments – on student confidence in a computer course. The results indicated that students reported higher levels of confidence on the Course Interest Survey (CIS) but not on the Instructional Materials Motivation Survey (IMMS). In a follow-up study, Huett et al. (2008) found that students who received confidence-building tactics did not demonstrate increased confidence. These results have interesting implications, as they suggest that targeting one construct within the ARCS-V model may be ineffective. Accordingly, the present study addressed all five ARCS-V constructs and a more holistic approach to a learning experience.

### *Ungrading and Confidence*

Alternative assessment strategies are one strategy to increase confidence, as the goal is to reinforce the idea that all students can succeed and that effort and learning are valued over performance. Ungrading is associated with increased confidence within empirical studies. In the literature, students enrolled in contract-graded, online, first-year composition courses reported that they were more confident that they could pass the course and that the instructor would value their best efforts (Stuckey et al., 2020). This increased confidence results in more experimentation, as research indicates that students enrolled in graded courses engage in less risk-taking and exhibit an aversion to failure (Gorichanaz, 2022; Guberman, 2021; McMorran et al., 2017; McMorran & Ragupathi, 2020). As Keller (2010) discussed, “Anxiety and fear are much greater parts of students’ lives than teachers realize” (p. 137). Grades can be a significant source of these emotions because students often are distracted by grades, viewing them as obstacles and endpoints (Harrison et al., 2015). Because the principal focus of ungrading pedagogy is extensive formative feedback in the absence of grades, ungrading may alleviate this distress, leading to increased confidence. Students also develop confidence when they attribute their successes to their own efforts (Jokelova, 2013). Because ungrading places effort at the center of the assessment process, it represents a promising strategy for increasing confidence.

As previously discussed, when students are more confident and less concerned about earning high grades, they are more likely to enroll in challenging courses (Chamberlin, 2018; McMorran et al., 2017; McMorran & Ragupathi, 2020; Schneider & Hutt, 2014). As McMorran and Ragupathi (2020) explained, students “often choose

majors and modules based not on their interests, but on what will lead to the highest GPA” (p. 927). Due to decreased confidence, Gorichanaz (2022) concluded that learners may try to “game the system” when grades are assigned, with students seeking to accumulate points rather than knowledge. These reactions are justified, as failure may result in substantial consequences in traditionally-graded courses. As Pulfrey et al. (2011) argued, “[F]ailure to succeed may not simply leave the individual at the starting block but worse off than before, with a lower GPA, loss of self-esteem and face, feelings of guilt, and potential retributions from powerful others” (p. 686). Rather than confidence to achieve a final grade, ungrading arguably encourages confidence behaviors, such as a growth mindset (see Dweck, 2016) and risk-taking because there is no punishment for making mistakes (Cowan, 2020; Ferguson, 2013; Gorichanaz, 2022; Guberman, 2021; Mallette & Hawks, 2020; McMorran et al., 2017; McMorran & Ragupathi, 2020). Inman and Powell (2018) highlighted that these concerns are especially relevant in composition courses because affect is a critical factor in “writing efficacy,” noting that many students “lack successful writing experiences, particularly writing experiences rewarded by grades” (p. 35). They further argued that “grades are insufficient in a process-based writing course” (Inman & Powell, 2018, p. 48). When students are not afraid to fail, they are more confident and more able to learn.

These issues relate to the aforementioned relevance construct in that students are more likely to choose courses that are personally meaningful when ungrading is employed. However, they also relate to confidence in that graded students often limit themselves to courses in which there is a high probability of success and where they are confident in their ability to earn a high grade. Ungrading seeks to subvert the cycle of



failure and subsequent decreased confidence by disturbing the signifying function of grades. Instead, as Stuckey et al. (2020) suggested, ungrading gives students “a clear path to success,” allowing them to focus on establishing their own goals (p. 3). Self-assessment is a core component of confidence-building in ungraded courses, as it increases confidence.

Angelo (2017) identified self-assessment as a key ARCS-V teaching strategy, arguing that students need instructor feedback but that “they also need self-assessment tools to help them become more aware of and better able to monitor, enhance, and maintain their own motivation to learn” (p. 103). When students can effectively monitor their own progress and evaluate their own writing, they become more confident in their ability to succeed. Jokelova (2013) also identified self-assessment as a key ARCS-V strategy, suggesting that students should be encouraged to evaluate their own work rather than relying on external evaluations. This type of self-assessment increases confidence. In alignment with the ARCS-V model, self-assessment also is an integral ungrading construct (Laflen & Sims, 2021). Ninomiya (2016) described that formative assessment represents a shift from convergent assessment, in which students are the recipients of assessment, to divergent assessment, in which students are initiators of assessment. This shift fosters confidence, encouraging students to become more self-reliant. Kohn (2011) distinguished this type of democratic self-assessment in which students take an active role in determining the assessment criteria from “the more common variant” in which students merely assess their progress toward the instructor’s predefined goals (p. 33). Although self-assessment is most clearly applicable to negotiated grades, it increases confidence with all forms of ungrading. Black et al. (2004) have suggested that students

should develop their own assessment criteria and create rubrics that they then use to grade their own work, claiming that active involvement in the process allows students to view themselves as “beneficiaries rather than the victims” of assessment (p. 16). Self-assessment is a vital part of this process because it helps students to develop the capacity for metacognition by encouraging them to analyze their own work (Black et al., 2004; Reinholz, 2016). Learning to trust themselves – including their ability to evaluate their writing, accurately identify writing weaknesses, and undertake actions to overcome these weaknesses – leads to increased self-efficacy and confidence.

### ***Satisfaction***

Satisfaction “refers to positive feelings about accomplishments and learning experiences” (Keller, 2017, p. 15). It is a critical component of continued motivation, as initial intrinsic motivation may be tempered by repeated failures or unfair teaching and grading practices (Keller, 2008b; Keller & Suzuki, 2004). Keller (2010) aligned intrinsic motivation with intrinsic satisfaction, which he characterizes as satisfaction “from feelings of mastery and from the pleasure of having succeeded at a task which was meaningful and challenging” (p. 166). It is preceded by novelty, competence, interest, and autonomy and requires sustained successful achievement (Keller, 2010). Keller (2010) identified required courses and grades as barriers to satisfaction. However, he also argued that “tangible extrinsic rewards,” such as grades, privileges, promotions, certificates, and tokens of achievement can provide satisfaction, provided that these rewards are distributed equitably and without favoritism (Keller, 2017, p. 15). For example, students must perceive that the amount of work is appropriate and that the assessments reflect learning objectives (Keller & Suzuki, 2004).

Keller (2010) identified three strategies for increasing satisfaction: intrinsic reinforcement, extrinsic rewards, and equity. Providing intrinsic reinforcement involves offering positive feedback. Providing extrinsic rewards involves incorporating rewards, including verbal praise, symbolic rewards, and incentives (Keller, 2010). Many strategies for increasing satisfaction fall under the broad heading of “positive recognition” (Keller, 2010, p. 190). For example, he recommended providing students with opportunities to use new skills, offering verbal reinforcement, and providing opportunities for students to teach new skills to others (Keller, 2010). Satisfaction also may be fostered through continuing motivation strategies, such as offering additional information about related subjects, informing students about ways to further pursue a topic of interest, and informing students about new areas to which the topic of interest applies (Keller, 2010). A review of the literature revealed no studies that have examined the satisfaction construct of the ARCS-V model in isolation. Thus, the role of satisfaction in the ARCS-V model represents a potential area for future research.

### ***Ungrading and Satisfaction***

Dissatisfaction can be especially pervasive in online courses because students typically do not form personal relationships with other students or with the instructor and because online learning requires great self-discipline, which many undergraduates do not possess (Ucar & Kumtepe, 2019). However, research shows that online student satisfaction increases when intrinsic and extrinsic motivational tactics are employed (Ucar & Kumtepe, 2019). In the ARCS-V model, satisfaction involves feelings of accomplishment and pride in one’s efforts (Keller, 2017). Ungrading arguably enhances satisfaction, in part, by increasing perceived competence (Pulfrey et al, 2011). When

students perceive that they can achieve learning outcomes, they are more intrinsically motivated and perform better, which fosters self-satisfaction and satisfaction with the course (Pulfrey et al., 2011). This facet of satisfaction has interesting implications in ungraded courses, as grades are removed as external rewards.

### ***Volition***

Keller (2008b) divided motivation into two levels. He characterized the first level as will, which includes the learner's desires, goals, and beliefs concerning the likelihood of success. The second level is volition, which is the process of taking definite actions to achieve identified goals (Keller et al., 2020). Thus, volition involves "the transition from desire, or motivation, to action especially when faced with competing goals" (Keller, 2020, p. 161). Because volition is necessary to sustain action, it is a precursor to persistence (Keller, 2017). As defined in the ARCS-V model, volition has two components: volition planning and volition control. Volition planning involves becoming attracted to a goal and committing to action, while volition control involves controlling actions and managing attitudes and behaviors in order to complete the goal (Keller, 2008a; Keller, 2008b; Keller, 2010). While motivation is primarily a mental process, volition is a demonstrable action (Keller, 2010).

### ***Ungrading and Volition***

Research suggests that traditional grades lead to diminished autonomy and agency, which are critical components of volition (Chamberlin et al., 2018). Indeed, Chamberlin et al. (2018) referred to autonomy as "perceived volition," arguing that grades "thwart autonomy" and that this lack of autonomy reduces task persistence (p. 2). In their seminal article on formative assessment, Hattie and Temperley (2007) maintained

that students at all levels of instruction conduct a cost-benefit analysis when determining whether to devote effort to improvement. When students perceive that the affective costs outweigh the potential benefits, they exhibit avoidance behaviors and, thus, reduced volition (Hattie & Timperley, 2007). In accordance with these findings, Pulfrey et al. (2011) found that grades lead to avoidance behaviors, arguing that “accomplishing a graded activity erodes autonomous motivation, and this erosion is responsible for the adoption of performance-avoidance goals” (p. 686).

Research has shown that grades undermine volition by creating a “transactional relationship between teacher and student” in which students have little agency over the learning and assessment process, becoming recipients rather than active agents who are responsible for their learning (Gorichanaz, 2022, p. 3). Gorichanaz (2022) further argued that grades create an “adversarial relationship” in which students “fight” for points by challenging and negotiating with the instructor (p. 5). Other researchers have claimed that the commodification of education has led to a “detached consumer-seller relationship” in which instructors award grades in exchange for tuition (Chamberlin et al., 2018, p. 8). These factors undermine trust and encourage powerlessness, dependence, and resentment (Chamberlin et al., 2018; Pulfrey et al., 2011). Conversely, increased agency and autonomy foster volition.

Although not explicitly discussed, there is data to describe the relationship between volition and ungrading. Students report experiencing more agency over their grades in ungraded courses (Lindemann & Harbke, 2011). Research suggests that ungrading redistributes power and control so that the instructor is a partner and a facilitator rather than a judge and a critic (Brubaker, 2010; Cowan, 2020; Ferguson, 2013;

Guberman, 2021; Hiller & Hietapelto, 2001; Litterio, 2016; Stuckey et al., 2020; Tannock, 2017). Participatory assessment, in which students are engaged in the assessment process, as is often the case with ungrading, was especially positively correlated with these affective benefits (Hawe & Parr, 2014; Ibarra-Saiz et al. 2020). Stuckey et al. (2020) found that ungrading promotes “student-centered pedagogical habits” and improved communication (p. 8). Other researchers have indirectly addressed volition. For example, Inman and Powell (2018) conducted a mixed-methods study in basic writing courses at a regional, research-intensive university in the United States. The study included 144 students enrolled in traditionally-graded courses and 219 students enrolled in labor-based contract graded courses (Inman & Powell, 2018). Following the study, they concluded that labor-based contract grading freed students from “the need for an authority figure’s mark of approval” (Inman & Powell, 2018, p. 41). Because increased autonomy and agency are associated with increased volition, these findings are especially relevant, as research suggests that learning is superficial when these concerns are ignored (Ninomiya, 2016). Thus, research indicates that ungrading is closely tied to volition.

### **Ungrading and Perceived Learning**

In addition to the constructs of the ARCS-V model, ungrading also is associated with increases in perceived learning. Indeed, research suggests that ungrading increases both actual and perceived learning indirectly by fostering a learning orientation rather than a performance orientation, as students learn more when they are focused on learning. Conversely, traditional grading may encourage a grade orientation rather than a learning orientation, which undermines the quality of learning by encouraging surface learning

techniques aimed at earning a high grade with learning as a secondary goal (Ferguson, 2013; Tippin et al., 2012). Research suggests that students perform better and demonstrate improved learning outcomes when ungrading is employed (Hiller and Hietapelto, 2001; Inman & Powell, 2018; Lindemann & Harbke, 2011).

Some studies have directly addressed the use of ungrading in writing classrooms. For example, Stuckey et al. (2020) conducted a mixed-methods study piloting the grading contract developed by Danielewicz and Elbow (2009) in half of students ( $N = 850$ ) enrolled in online, asynchronous, accelerated first-year composition courses at Arizona State University. At the end of the term, 53 students from traditionally-graded sections and 44 students in the contract graded sections completed a survey designed to elicit student opinions regarding the grading method employed in the course. The results showed that students in the contract graded sections rated their level of perceived learning significantly higher than those in traditionally-graded courses.

Research also suggests that students are better able to analyze and revise their writing when ungrading is employed. For example, Chen and Zhang (2017) conducted a mixed-methods study in EFL composition courses at a Chinese university. Students were assigned to either a summative assessment ( $N = 30$ ) or formative assessment ( $N = 27$ ) group based on section enrollment (Chen & Zhang, 2017). The results indicated that improved actual learning, as students in the formative assessment sections demonstrated improvements in essay content, language, structure, and academic format. Finally, ungrading is associated with long-term learning gains and improved knowledge retention. Following a four-year implementation of individualized, negotiated grading contracts with 473 students in 22 undergraduate and graduate management courses at three

universities in the United States, Hiller and Hietapelto (2001) concluded that contract grading increases actual learning and work quality.

### **Theoretical and Pedagogical Limitations of Ungrading**

Although the literature search revealed no empirical studies that directly contradict the benefits of ungrading, some researchers have challenged ungrading on theoretical grounds, suggesting that ungrading solves some problems while exacerbating others. Many of the counterclaims can be described as issues related to grade integrity. The primary challenges and limitations of ungrading as represented in the literature are described below.

#### ***Complacency***

The most common criticism of ungrading is that it leads to complacency, both among students and instructors. McMorran and Ragupathi (2020) conducted a study at a public university in Singapore where a grade-free first semester program had been implemented. During the first 18 months of the program, 3,291 students and faculty completed surveys about their experiences (McMorran & Ragupathi, 2020). Ninety-eight percent of the responses were from students (McMorran & Ragupathi, 2020). An additional survey was distributed to 498 faculty. Although both students and faculty identified many benefits associated with the grade-free first semester, some instructors expressed concerns that ungrading resulted in lower standards or teaching aimed at the weakest students (McMorran & Ragupathi, 2020). Conversely, there is some evidence that ungrading may negatively affect relevance. Some research also indicates that ungrading may perpetuate the problems it is intended to solve. For example, McMorran



et al. (2017) found that it can be difficult to convince students that learning itself is meaningful.

Researchers also have argued that grades are essential precursors of motivation and that ungrading results in other aspects of complacency, such as reduced participation. For example, Merva (2003) noted that faculty using pass-fail ungrading reported that students deemphasized coursework under a pass/fail grading system and that classroom discussions suffered due to lack of preparation by students. She further argued that some students viewed the course as an “academic vacation” (p. 155). Additionally, some studies have found that instructors perceived that ungrading decreased motivation (McMorran & Ragupathi, 2020). Similarly, in a study of student attitudes about pass-fail grading among graduate law school and education students at a mid-sized, highly selective university in the United States, Michaelides and Kirshner (2005) found that 79% students indicated that they exerted more effort in graded courses. This problem can be more pronounced at open enrollment colleges and universities, where there are fewer highly motivated high achievers. These findings diverge from the mass of ungrading research, which indicates that ungrading increases motivation. However, they suggest that the role of motivation in ungrading is still unclear, especially given the qualitative nature of the data. Additionally, more studies are needed to determine the complex relationship between grades and complacency.

The counterargument is that, while grades are motivating, research consistently indicates that grades foster extrinsic motivation while inhibiting intrinsic motivation (Chamberlin et al., 2018; Gorichanaz, 2022; Pulfrey et al., 2011; White & Fantone, 2010). This distinction is critical because intrinsic motivation is essential for a variety of

affective learning gains (e.g., self-efficacy, self-determination, and life-long learning). As McMorran & Ragupathi (2020) explained, “[S]tudies show the difficulty of developing intrinsic reasons for learning, such as for the joy of learning or the development of skills, when students are preoccupied with grades” (p. 362). Researchers also have noted that grades engender indifference because they do not promote course interest or deep learning (Chamberlin et al., 2018; Gorichanaz, 2022). Lack of interest and superficial learning may engender complacency. Instead, research indicates that traditional grading results in a complacent surface learning approach in which students only study the information that is likely to be included in exams and that they quickly forget this information once the exam has been completed (Tippen et al., 2012) These issues warrant further study into motivation and ungrading, a gap which the study sought to address.

### ***Grade Inflation***

The second most common concern is that ungrading inhibits grade integrity by fostering grade inflation. However, grade inflation is already pervasive in the absence of ungrading, and there is no indication that ungrading exacerbates grade inflation. Grade inflation is real. Research does indicate a gradual upward shift in average grades over time. Poorman and Mastorovich (2018) related that the “current average grade in colleges is an A, which represents a 12% increase since 1988” (p. 322). Guberman (2021) further explained that there has been a steady rise in average grades since 1990, noting that no studies have directly examined the reasons for this phenomenon. However, it is difficult to determine whether grade inflation is attributable to changes in grading practices or whether the change reflects actual improved performance. Indeed, there are myriad potential reasons for grade inflation that may be unrelated to grading practices.

Research into the relationship between ungrading and grade inflation has been mixed. Lindemann and Harbke (2011) found that students were three-times more likely to earn a grade of A in their contracted-graded course. However, other researchers have obtained different results. For example, in a study of 188 enrolled in nine sections of undergraduate English courses, Potts (2010) found that most contract-graded students earned a grade that was within five points of the grade they would have earned had the course been graded using traditional methods, concluding that grading contracts are “impressively accurate” (p. 35). Similarly, Guberman (2021) found that fewer than 10 percent of 50 students enrolled in online, undergraduate history courses proposed a grade that differed from the grade proposed by the instructor when negotiated grading was employed. Many students proposed a grade that was lower than the grade proposed by the instructor (Guberman, 2021). Stuckey et al. (2020) found that while approximately 50 percent of students in both the graded and ungraded sections earned a grade of A or B, the number of students who earned an A was significantly lower in the contract graded sections. These mixed results make it difficult to draw any decisive conclusions about ungrading and grade inflation. Nonetheless, as Sadler (2009) argued, concerns about grade inflation reflect a “deep-seated belief that grade integrity matters” (p. 823). This issue warrants further research, and future studies could examine the role of ungrading practices in perpetuating grade inflation.

### ***Measurement Theory***

Next, many measurement theorists challenge ungrading from a methodological standpoint, primarily from the perspective of classical test theory, arguing that ungrading prohibits grade integrity. Fundamentally, grades are a form of measurement. As Landrum

and Dietz (2006) asserted, all forms of assessment are “prone to multiple types of measurement error” (p. 299). As such, they are comprised of a true score and error. Rather than adopting an ungrading approach, these scholars instead argue that the goal should be to minimize error so that grades reflect true scores (Allen, 2005; Sadler, 2009). Additionally, these researchers contend that grades should conform with the principles that apply to all measurements, such as adequate validity and reliability (Allen, 2005; Iamarino, 2014, Sadler, 2009; Yang et al., 2015). For example, Sadler (2009) asserted that credible, valid, and reliable grades preclude the incorporation of factors such as participation, effort, and improvement. Because many forms of ungrading are founded in labor and effort, these methods of assessment often violate the fundamental assumptions of measurement theory and the traditional purposes of grading. Theorists have proposed a variety of methods for improving the reliability and validity of grades. For example, many researchers have argued that nonacademic factors should be reported separately from grades (Allen, 2005; Iamarino, 2014; Sadler, 2009). Rubrics are another way to make grades more valid and reliable (Brookhart et al., 2016; Schinske & Tanner, 2014; Yang et al., 2015). Finally, instructors can be trained in sound assessment practices (Allen, 2005; Brookhart et al., 2016; Yang et al., 2015). As Yang et al. (2015) related, “[I]nterrater reliability tends to be high if the performance tasks are the same across all students, raters are well-trained, and scoring rubrics are well established” (p. 467). Thus, research suggests that ungrading may violate the tenets of classical test theory. However, even in ungraded courses, educators may take actions to mitigate this problem.

## ***Grade Reform***

Many theorists also have acknowledged the problems with traditional grading while arguing that the solution to the problems with a flawed assessment systems does not necessitate a wholesale abdication of grades, proposing grade reform rather than grade rejection. Many of these arguments propose standards-based grading (SBG) as a solution (Brookhart et al., 2016; Iamarino, 2014; Sadler, 2009). SBG involves comparing student achievement against “fixed external anchor points” that are determined in advance by educational institutions and the academy as a whole (Sadler, 2009). Iamarino (2014) argued that SBG is especially applicable to domains like writing because of its focus on formative assessment. Thus, SBG shares similarities with ungrading. First, formative feedback is an important component of SBG (Iamarino, 2014). Additionally, it involves a collapse of the traditional letter grade system to fewer gradations, such as *met*, *approaching*, and *unmet* (Iamarino, 2014). However, SBG differs from ungrading in that performance is judged in reference to specific, pre-determined external criteria rather than grading rubrics or policies constructed by individual instructors or departments. SBG also is detached from nonacademic factors, such as effort and participation (Iamarino, 2014; Sadler, 2009). Although SBG is touted as a more valid and reliable assessment method, some research indicated that it is subject to the same problems that accompany all forms of assessment (e.g., Cizek, 2000). Indeed, both research and practice indicate that no perfect solution exists.

## ***Ethics***

Some researchers view grade integrity as an ethical issue. Sadler (2009), a prominent expert on assessment, argued that grades should be “strictly commensurate

with the quality, breadth, and depth of students' academic achievement" (p. 807). Similarly, Allen (2005) argued that the sole purpose of grades is to accurately communicate a student's level of academic achievement. Because grades are used to make admissions, employment, scholarship, and accreditation decisions, they have real and lasting consequences for students and organizations. These consequences are particularly relevant for high achieving students, who rely on grades to differentiate them from less able peers (Reddan, 2013). Additionally, admissions and hiring agents interpret grades as signifiers of achievement, as they represent a "common currency" among and between educational institutions and industry (Sadler, 2009, p. 810). These consequences present an ethical obligation to ensure that grades represent actual achievement, as grades fail to fulfill their intended purpose when external factors are incorporated such that comparisons among students are no longer credible. Researchers also argue that incorporating nonacademic factors into grades sends a message to students that merely completing work or turning it in on time is more important than meeting learning objectives (Iamarino, 2014). This misconception also poses an ethical issue because it provides students with inaccurate information about their level of learning and because it perpetuates confusion about the fundamental purposes of assessment.

### ***Student Preferences***

Research indicates that some students prefer traditional grades over ungrading methods. For example, in a reflection on their experiences with labor-based contract ungrading in technical communication courses, Mallette and Hawks (2020) expressed that some students viewed grading contracts as unclear and preferred the structure provided by a traditional grading system. Similarly, in a narrative concerning their

experiences with contract grading in a summer bridge program at College of the Holy Cross, Reardon and Guardado-Menjivar (2020) found that some students in their disliked ungrading because they felt unable to gauge their performance, preferring grades on homework assignments and essays so that they could track their progress throughout the semester. Reardon and Guardado-Menjivar (2020) further observed that some students were unable or unwilling to discard grades, explaining that grades still permeated their conversations with students. Inman and Powell (2018) similarly observed that grades had deep-seated meanings for students, explaining that “their very identities and their progress to the identities they desire are enmeshed with grades in ways that contract grading did not seem to address for them” (p. 41). Research similarly indicates that many students prefer grades simply because they are familiar (Brilleslyper et al., 2012; Ferguson, 2016; Inman & Powell, 2018; Reardon and Guardado-Menjivar, 2020). As Cowan (2020) has explained, “Grades are a language that students have been taught to speak” (p. 6). Most students have participated in traditionally-graded courses throughout their educational experiences, and ungrading can be disorienting and uncomfortable for them, especially at the beginning of the semester before any assignments have been evaluated (Brilleslyper et al., 2012; Ferguson, 2016; Spidell & Thelin, 2006).

No form of assessment works equally well for all students or in all courses (Cowan, 2020). Some researchers have attempted to quantify student approval ratings for ungrading. As previously discussed, McMorran and Ragupathi (2020) found that 66% of students approved of ungrading during their first semester, while 72% approved of ungrading during their third semester. They noted that ungrading “fostered the simultaneous reduction and creation of stress” with some students, often high achievers,

preferring grades (p. 373). Similarly, Hiller and Hietapelto (2001) reported that 94% of their students preferred ungrading. In a survey of 71 economics students at a university in Sweden, Nystrom (2018) found that 56.3% of students preferred pass-fail ungrading, while 43.7% preferred traditional letter grading. However, they noted that students who perceived a high probability of earning an A were more likely to choose letter grades (Nystrom, 2018). In a study of 90 undergraduate journalism students at a university in Canada, Barber (2021) also found that students expressed a strong preference for traditional grades. Ferguson (2016) similarly noted that high achieving students preferred traditional letter grades and that lower performing students preferred ungrading methods. Students also questioned the fairness of ungrading. For example, Spidell and Thelin (2006) found that higher performing students resented that lower performing students were provided multiple revision opportunities. Collectively, the research indicates that student grading preferences are quite diverse. However, the research does suggest that most students prefer ungrading to traditional assessment.

### ***Ungrading is Grading***

Finally, some researchers and scholars have argued that ungrading is merely traditional grading with different terminology. Because most universities require instructors to assign final grades, university policies typically dictate that ungrading ends with a grade. For example, McMorran and Ragupathi (2020) found that some students in their study viewed ungrading as “an irresponsible delay of the inevitable: learning that will be graded” (p. 932). Ungrading is a form of assessment. However, all forms of ungrading differ from traditional grading in fundamental ways. Most importantly, ungrading values effort and labor rather than achievement. As Smith (1999) noted,



ungrading rewards students who are “unspectacular but diligent in their coursework” (p. 430). Traditional grades are summative designations of achievement, which are then permanently inscribed on a transcript. By contrast, *accept* and *revise* are interim designations provided on individual assignments, which may then be revised until the student reaches a minimum level of competence. In this way, students in ungraded courses are measured against their own prior performance rather than adherence to external standards, as in criterion-referenced and standards-based grading, or comparison with other students, as in norm-referenced grading. Ungrading also differs from these grading systems in that summative assessment is eschewed in favor of extensive formative feedback. Minimal competence earns students a C under the traditional letter grade scale with grades of A or B requiring students to exhibit either excellent or above average achievement, respectively. With the notable exception of the unilateral grading contract proposed by Danielewicz and Elbow (2009), in which students earn a B for minimal competence, students typically earn an A for completing assignments at the minimum level of competence in ungraded courses. Lower grades are assigned only when students fail to complete assignments.

### **Conclusion**

Developing instruction that promotes motivation (attention, relevance, confidence, and satisfaction), volition, and perceived learning is a critical facet of effective instructional design. Research suggests that ungrading may be a potential strategy for achieving these aims. The benefits of ungrading are well-represented in the literature. However, many valid arguments against the use of ungrading exist simultaneously with the potential benefits. In many ways, ungrading represents a

departure from traditional conceptions of the fundamental purposes of grades. Theorists who view grades as academic currency that must accurately and reliably communicate achievement are unlikely to favor ungrading. Conversely, those who view grades as fundamentally flawed measures associated with motivational, volitional, and learning deficits are unlikely to be swayed by arguments situated in classical assessment theory. In practice, the various ungrading methods fall on a continuum between measuring achievement and measuring effort and labor. Each gradation on this continuum reflects real and fundamental ideologies. This tension suggests a need for empirical evidence into the effectiveness and consequences of grading methods. Although ungrading has not been studied in the context of the ARCS-V motivational model, the model is well-suited to research addressing motivational problems in educational settings, and this gap represents a potentially fruitful avenue for future research. The study sought to fill this gap by drawing connections between the five constructs of the ARCS-V model and ungrading pedagogy.

## **CHAPTER THREE: METHODOLOGY**

### **Introduction**

Lack of motivation (attention, relevance, confidence, and satisfaction), volition, and perceived learning are critical challenges in undergraduate research and argumentative writing (RAW) courses at the university of interest, suggesting a need for innovative instructional strategies aimed at increasing these attributes. Ungrading has been proposed as a potential solution to these problems. Accordingly, the study examined how student motivation (attention, relevance, confidence, and satisfaction), volition, and perceived learning change over time when ungrading is used in RAW courses. This chapter describes the methodology for the study, beginning with an investigation plan that outlines the basic methodological structure of the study. This section is followed by a description of the participants, setting, instrumentation, and procedures. Finally, the chapter provides a detailed description of the data analysis phase of the study. In addition to detailing the methodology for the study, this section should allow future researchers to replicate the study in other settings.

### **The Investigation Plan**

A quantitative, survey-based, repeated measures, correlational design was used in the study. During the spring 2023 semester, 57 students enrolled in seven online sections of RAW participated in an ungraded version of the course. Students self-enrolled in the course, so random sampling was not possible. Therefore, convenience sampling was used. The university of interest uses a 16-week semester. Students completed the surveys during week one (T<sub>1</sub>), week four (T<sub>2</sub>), and week eight (T<sub>3</sub>) of the course. This schedule was chosen because the first major essay is due during week four, and the second major

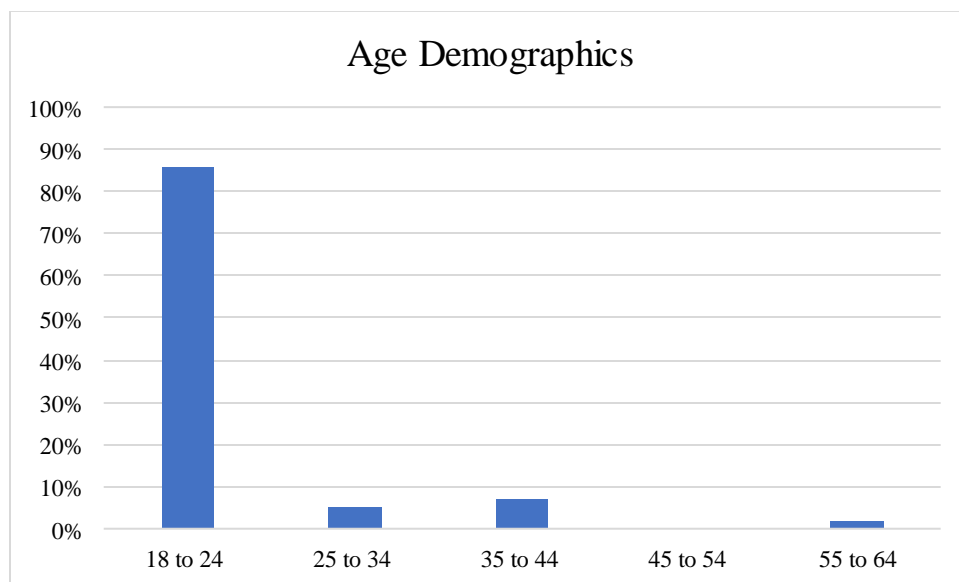
essay is due during week eight. Thus, students were anticipating feedback for the first essay when the instruments were administered at T<sub>1</sub>, had received feedback for the first major essay at T<sub>2</sub>, and were anticipating feedback for the second essay when the instruments were administered at T<sub>3</sub>. Another reason for the proposed schedule was that the time between surveys was equidistant. As data for the study was quantitative, descriptive and inferential statistics were used to analyze the data and test the null hypotheses. The data collection and analysis procedures are outlined in detail below.

### **Participants**

The accessible population for the study included all undergraduate students enrolled in online RAW courses at the university. During the fall 2022 semester, approximately 800 students were enrolled in undergraduate RAW courses with approximately 250 students enrolled in online sections of the course. For the present study, this population was narrowed to a non-randomized convenience sample of students enrolled in seven online sections of RAW taught by two non-tenure-track instructors during the spring 2023 semester. Both instructors had been using ungrading for approximately one year. Students' prior experience with ungrading was not collected in the demographic survey. However, ungrading is a relatively uncommon practice, and it is likely that student familiarity with ungrading was low. A total of 118 students were enrolled in the seven online course sections. Students were able to opt out of the study via the informed consent form. Eighty-seven students agreed to participate in the study. Of these, 30 students failed to complete one or more surveys. The final sample was comprised of 57 students. Thus, the sample for the study represented approximately 23% of the accessible population. Twenty-nine participants (51%) were enrolled in the

primary researcher's courses, and 28 were (49%) enrolled in courses taught by a colleague. An a priori power analysis for the study was conducted using G\*Power 3.1.9.4 to determine the minimum sample size required to test the null hypotheses. The results indicated that a sample size of 61 was required to achieve 80% power for detecting a medium effect with an  $\alpha = 0.05$  significance criterion for a within-group, repeated-measures, analysis of variance (ANOVA). However, it was determined that the loss of statistical power was marginal and that the sample could be used to test the study hypotheses.

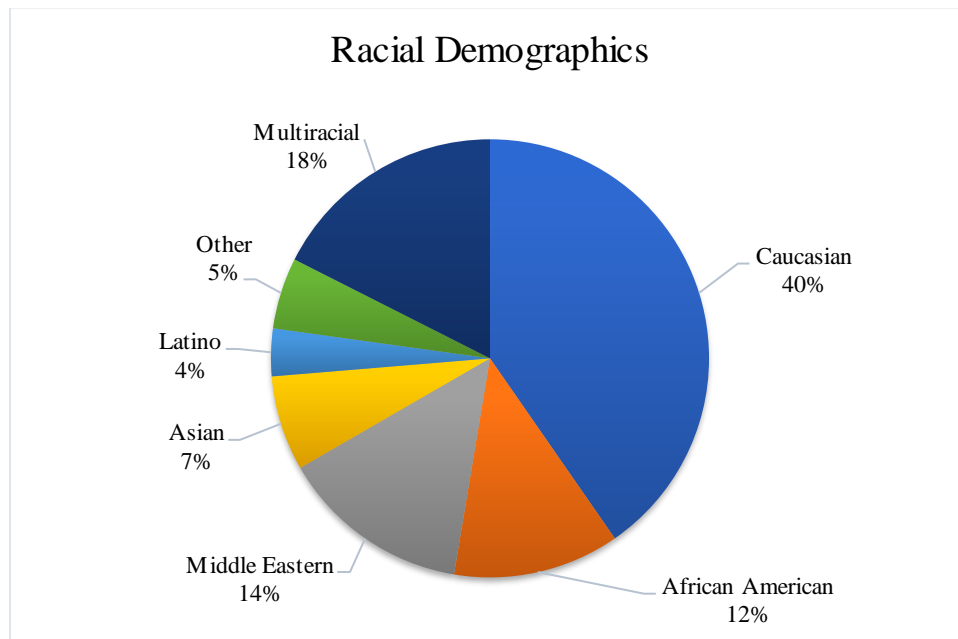
To obtain approximate demographic data for the study sample, a demographic survey was distributed to participants during the first week of the semester. The demographic characteristics of the sample were somewhat homogeneous. Twenty-eight participants (49%) were male, and 28 (49%) were female. One participant identified as Other (2%). Their ages ranged from 18 to 56 years with a mean age of 21 years. However, as represented in Figure 2, most participants (84%) were between the ages of 18 and 24.



## Figure 2

### *Sample Age Demographics*

The ethnicities of the participants were somewhat diverse. Twenty-three participants (40%) identified as Caucasian, 10 (18%) as Multiracial, 8 (14%) as Middle Eastern, 7 (12%) as African American, 4 (7%) as Asian, 2 (4%) as Latino, and 3 (5%) as Other. Figure 3 summarizes the racial demographics of the sample.



## Figure 3

### *Sample Racial Demographics*

A rather narrow range of socioeconomic statuses was represented. Most students were either working class ( $n = 27$ ) or middle class ( $n = 27$ ). Three participants identified as poor, and no participants identified as affluent. Most participants were currently employed either full-time ( $n = 13$ ) or part-time ( $n = 32$ ). Twelve participants were not employed. Although most participants were first-year students ( $n = 31$ ), there were 16 sophomores and 10 juniors. All participants had completed the equivalent of one

semester of expository writing prior to enrolling in the study, as it is a prerequisite for the RAW course. However, some students enroll in the course during their first semester, having completed expository writing as dual-enrollment high school students or having tested out of expository writing through American College Testing (ACT) or College Level Examination Program (CLEP) scores.

### **Setting**

The setting for the study was the English department at a large state university in a mid-size city in the Southeastern United States. Approximately 19,000 undergraduate students are enrolled each semester. This site was chosen because the researcher is an instructor in the department. The English department is the largest on campus with approximately 80 full-time faculty who serve over 5,000 general education students each semester. Because university policy requires that online courses be fully-asynchronous, with no required class meetings, all instructional interventions and communication was technology mediated.

### **Intervention**

The study included one intervention. All participants completed an ungraded version of the course. All other aspects of the course were held constant or statistically controlled. Because ungrading is an umbrella term for a variety of pedagogical strategies, it is important to establish the definition of ungrading that was used in the study. Because university policy requires instructors to assign final grades, narrative grading (i.e., no letter or number grades at all) was not an option. Therefore, for the present study, the instructors used labor-based grading in which the instructor assigns a grade based on the number of assignments completed. There were no numerical or letter grades on

individual assignments. Instead, assignments were marked as *accept* or *revise*, and narrative feedback was provided. Final grades in the course are assigned based on the following specifications:

- Grade of A: Completion of three essays and an annotated bibliography.  
Completion of at least 30 out of 35 discussion posts and homework assignments.
- Grade of B: Completion of three essays and an annotated bibliography.  
Completion of at least 25 out of 35 discussion posts and homework assignments.
- Grade of C: Completion of three essays and an annotated bibliography.  
Completion of at least 20 out of 35 discussion posts and homework assignments.
- Grade of D: Failure to complete one or more essays and/or the annotated bibliography. Completion of at least 15 out of 35 discussion posts and homework assignments.
- Grade of F: Failure to complete one or more essays and/or the annotated bibliography. Completion of fewer than 15 of the 35 discussion posts and homework assignments.

In alignment with departmental policy, students must complete all four major projects (i.e., three essays and the annotated bibliography) in order to earn a passing grade for the course. This “bundling” of assignments, in which students choose to complete a specified number of tasks for each letter grade, aligns with ungrading approaches used in other ungraded courses, especially when labor-based ungrading is used (e.g., Guberman, 2021; Reardon & Guardado-Menjivar, 2020; Reichert, 2003; Smith, 1999).

Students in all sections of the course complete four major projects: three 1,000-word argumentative research essays and an annotated bibliography comprised of eight



scholarly sources. The university uses a 16-week semester system, and students have one month to complete each major project. Assignments are due on Sunday, Tuesday, and Thursday of each week. All RAW instructors at the university use a process-based approach to writing, and students complete informal process assignments throughout the course, including a proposal, a rough draft, and a peer review for each major project. Additionally, students complete weekly discussion posts, which invite them to summarize and reflect on the reading assignments. Students are required to write a 200-word post and to respond to posts written by two other students. Process assignments and discussion posts are graded based on completion (i.e., students receive full credit for completing them), although narrative feedback is provided. The class is heavily textbook-based, as all assigned readings are contained in *They Say/I Say: The Moves That Matter in Academic Writing with Readings, 4th ed.* Descriptions of the four major projects are included in Appendices B-E. Apart from the textbook readings, all course content is contained in the Desire2Learn (D2L) learning management system. With the exception of rough drafts of the major projects, all assignments are submitted via the D2L dropbox function. Students post rough drafts of each major project to the D2L discussion board for peer review.

### **Instrumentation**

Three instruments were used in the study. Motivation was measured by the Course Interest Survey (CIS), and volition was measured by the Volition for Learning Scale (VFLS). Both instruments were developed for the purpose of measuring ARCS constructs, and both have been validated for internal consistency (Keller, 2010; Keller et al., 2020). Perceived learning was measured using the CAP Perceived Learning Scale

(CAP Scale). It has been validated (Rovai et al., 2009) and is widely used in educational research (e.g., Akbaş et al, 2016; Rockinson-Szapkiw et al., 2013; Wighting et al., 2011).

The three instruments are provided in Appendices F-H.

### **Motivation**

The CIS was used to measure motivation. The scale was designed by Keller (2010) and is aligned with the motivation construct in the ARCS-V model. It is one of two instruments used to measure motivation in the model: the Instructional Materials Motivation Survey (IMMS) and the CIS. The CIS was appropriate for the study, as it is used to measure motivation in instructor-led environments, whereas the IMMS measures motivation towards instructional materials in self-directed learning environments (Keller, 2010). Although the study was limited to online courses, these courses are not entirely self-directed, as students do interact with the instructor through discussion boards, emails, and videos. Additionally, the study examined motivation at the course level, which is more amenable to the CIS. Keller (2010) explained that the CIS was not designed to measure academic motivation generally. Instead, it is used to measure student motivation towards a specific course.

The CIS was developed by creating a pool of items for each of the four original ARCS constructs (Keller, 2010). Ten graduate students “who were well versed in the motivational literature” then reviewed these items and discussed those that seemed ambiguous or unrelated to the four constructs of the ARCS model (Keller, 2010, p. 278). Keller (2010) did not provide demographic information for these students. Next, a second group of 10 individuals, who were mostly graduate students but not experts in the learning motivation literature, were asked to answer each question in the revised scale

twice, once as a highly motivated student and then as an unmotivated student. Again, Keller (2010) did not provide demographic information for the second group of students. Based on the results of this analysis, the scale was revised. This analysis revealed a few ambiguous items that were “poor discriminators” (Keller, 2010, p. 278). These items were revised or deleted.

The final questionnaire is comprised of 34 five-point Likert-type items. Each item is presented as a declarative statement, and respondents are asked about their level of agreement or disagreement with each statement. Scores for individual items range from 1, *not true*, to 5, *very true*. The items are divided by construct into four subscales. The attention subscale has eight items; the relevance subscale has nine items; the confidence subscale has eight items, and the satisfaction subscale has nine items. Cumulative scores for the CIS are calculated by adding the scores on the subscales. Thus, the minimum score is 34, and the maximum score is 170. Nine items are reverse scored. It is designed to be administered by instructors as a self-report questionnaire, either online or on paper.

Keller (2010) evaluated reliability for each construct by administering the scale to 45 undergraduate students at a large state university in the Southeast. Keller (2010) did not provide the sample demographics. A pretest version of the scale was created by changing the verb tenses of the items to future tense (Keller, 2010). This version of the scale was administered to 65 undergraduate students (Keller, 2010). Following the administration of the pretest version, some items were revised. Finally, the standard version of the scale was administered to 200 undergraduate and graduate students in the School of Education (Keller, 2010). The Cronbach’s coefficient alpha (Cronbach’s  $\alpha$ ) reliability estimates for each subscale were as follows: attention .84, relevance .84,

confidence .81, satisfaction .88, and total scale .95 (Keller, 2010). Thus, the internal consistency was judged as satisfactory. Test-retest reliability was not reported. Other studies have produced similar results (e.g., Gabrielle, 2003; Huett, 2006).

Validity is a central concern when selecting instruments for any study. In recent years, researchers have departed from a “tripartite” concept of validity, comprised of content, criterion-related, and construct validity (Bandalos, 2018). Instead, the *Standards for Educational and Psychological Measurement* now direct researchers to consider five sources of validity evidence when evaluating the uses and consequences of tests (Bandalos, 2018). The sources of evidence include evidence based on test content, response processes, internal structure, relations to other variables, and the consequences of testing (Bandalos, 2018).

In validating the CIS, Keller (2010) argued that because the four subscales are highly intercorrelated and because the instruments were developed to measure “situation-specific attitudes,” it is “difficult to apply traditional factor analysis to the instrument and obtain this factor structure,” explaining that “other methods were used to support the conceptual structure” of the measure (p. 286). Keller (2010) described the validation process for the CIS as follows. Course grades and grade point averages for the 200 students who participated in the reliability study were correlated with scores on the CIS. Keller (2010) argued that because these correlations were significant at the 0.05 level, the study “supports the validity of the CIS as a situation-specific measure of motivation” (Keller, 2010). However, this initial validation did not use robust statistical analyses or traditional validation methods.

An extensive literature search yielded only one additional validation study. Naveed and Bhatti (2017) validated the instrument in a quantitative, cross-sectional study using a simple random sample of 220 students at a women's medical college in Pakistan. The researchers did not state whether the English version or a translated version of the instrument was used. All participants were female and were between the ages of eighteen and twenty-four (Naveed & Bhatti, 2017). Although specific demographic data was not provided, the researchers noted that the participants were ethnically and culturally diverse (Naveed & Bhatti, 2017). All participants completed the CIS at the end of their first term (Naveed & Bhatti, 2017). Reliability analysis showed a Cronbach's  $\alpha$  of .78 for the attention construct, .70 for the relevance construct, .69 for the satisfaction construct, and .86 for the total scale (Naveed & Bhatti, 2017). However, the reliability of the confidence construct was .24, which was judged as unacceptably low (Naveed & Bhatti, 2017). The researchers found that reliability of this subscale would improve if items 6, 7, and 17 were deleted (Naveed & Bhatti, 2017). However, they also noted that the low reliability may have resulted from "random error or difficulty in item interpretation" or from the fact that the study was conducted in a different culture and context than Keller's (2010) study (Naveed & Bhatti, 2017). Because the reliability of the other three constructs was acceptable and because the reliability of the total scale was .86, they concluded that the reliability of the CIS was acceptable (Naveed & Bhatti, 2017).

Principal component analysis (PCA) was used to measure the latent structure of the CIS, providing validity evidence based on internal structure (Naveed & Bhatti, 2017). This analysis revealed a five-factor model, as five components had eigenvalues greater than one (Naveed & Bhatti, 2017). The five-factor model accounted for 55% of the

variance (Naveed & Bhatti, 2017). An examination of the scree plot indicated a four-factor solution, and a Varimax orthogonal rotation also suggested a four-factor model with four components accounting for 51% of the variance (Naveed & Bhatti, 2017). The four subscales were highly correlated with each other, which the researchers accepted as evidence that the four subscales measure different facets of the same construct (Naveed & Bhatti, 2017). The researchers concluded that the “CIS has a valid four factor structure for measuring motivation” in the study context (Naveed & Bhatti, 2017). The study did not specifically evaluate validity evidence based on test content, response processes, relations to other variables, or consequences of testing.

Several characteristics of the CIS must be considered when administering and scoring the instrument. Because the CIS is designed to be administered as a pretest and posttest, the verb tenses of the items can be altered to reflect the temporal context (Keller, 2010). Additionally, because the number of items varies among subscales, Keller (2010) recommends calculating a mean score for each subscale, which allows for direct comparison between subscales. Another important point is that scores on the CIS cannot be designated as high or low, as the scale is not norm-referenced (Keller, 2010). One can only compare scores obtained at one time with scores at a subsequent time (Keller, 2010). Thus, the scale is designed to be used as a pretest and a posttest (Keller, 2010).

### **Volition**

The VFSL was used to measure volition. Although this scale is rather new, having been developed in 2020, it was chosen for this study because it was designed by Keller et al. (2020) to measure volition using the constructs of the ARCS-V model and because it was developed and validated in online courses. As Keller et al. (2020) have argued, the

existing scales for measuring motivation “are important and have value and expediency in their context,” but none measure this construct as defined by the ARCS-V model (p. 164). The VFLS was developed through a rigorous process of continual refinement. First, a literature review and a content analysis of existing instruments for measuring volition resulted in an initial pool of 64 items. Three experts “who had carried out research on motivation and volition” then assessed this set of items on a 4-point Likert-type scale ranging from 1, *not representative*, to 4, *strongly representative*, of the construct that each item was designed to measure (Keller et al., 2020, p. 186). Refinement decisions were determined by the conformity of two or more experts on the 4-point scale. This process resulted in a narrowed pool of 57 items. The scale was further refined after being reviewed by a second set of four experts with experience in scale development, who assessed the items for ambiguity. After revising the items for clarity, the scale was reviewed by 13 undergraduate seniors who had completed online courses in the past using a 5-point Likert-type scale, ranging from 1, *completely disagree*, to 5, *completely agree*. Keller et al. (2020) did not provide the demographics of this group of students. The researchers then further revised the items to reflect student feedback. After calculating item-total correlations, skewness, and kurtosis values, the scale was reduced to 39 items by deleting items with correlations of less than 0.4. Using exploratory factor analysis (EFA), the scale was limited to 13 items with two factors: volition planning and volition control (Keller et al., 2020). The final VFLS contains 13 five-point Likert-type items. Each item is presented as a declarative statement, and respondents are asked about their level of agreement or disagreement with each statement. Scores on individual items range from 1, *completely disagree*, to 5, *completely agree*. The two subscales are

intended to be administered together, with items 1-5 measuring volition planning and items 6-13 measuring volition control (Keller et al., 2020). Thus, total scores on the VFLS range from 13 to 65. No items are reverse scored. The scale is designed to be administered by instructors or instructional designers as a self-report questionnaire, either online or on paper (Keller et al., 2020).

Keller et al. (2020), assessed the reliability of the scale by administering it to a convenience sample 594 students enrolled in online English courses taught by five faculty at a state university in Turkey. The sample was comprised of 311 females and 283 males. Of these, 508 were in the 18-20 age range, and 86 were over 21. Reliability was measured using Cronbach's  $\alpha$ . The reliability estimates were .732 for the first subscale and .809 for the second subscale with a total Cronbach's  $\alpha$  of .825 (Keller et al., 2020). Composite reliability was between .77 and .84. Test-retest reliability was not reported.

Keller et al. (2020) also assessed the validity of the scale. Validity evidence based on internal structure was assessed by performing a CFA. The 13-item, two-factor structure was found to be a good fit for the data. After determining that the CFA model was acceptable, the convergent validity was evaluated by controlling factor values of .40 and above. Average variance was approximately .40, which is below the .50 threshold generally accepted in the educational literature (Keller et al., 2020). However, the researchers asserted that a .40 average variance is acceptable if the composite reliability is greater than .60 and judged the convergent validity to be acceptable (Keller et al., 2020). Apart from addressing validity based on test content during the scale development phase, Keller et al. (2020) did not evaluate validity evidence based on content, response processes, relations to other variables, or consequences of testing.



## **Perceived Learning**

Perceived learning was measured using the CAP Scale. Although other instruments for measuring perceived learning exist, the CAP Scale was selected because it measures learning in three domains: cognitive, affective, and psychomotor (Rovai et al., 2009). This contrasts with measures of perceived learning that address the cognitive domain only. Additionally, previous perceived learning scales were developed before distance learning became prevalent and are, therefore, designed to be used in face-to-face (F2F) courses (e.g., McCroskey, 1994; Richmond et al., 1987), whereas the CAP Scale was assessed for reliability and validity evidence in both F2F and online courses. Indeed, Rovai et al. (2009) argued, “One of the immediate benefits of the CAP Scale is its potential use within online learning research” (p. 11). The CAP Scale also is more comprehensive than other perceived learning scales. For example, prior to the development of the CAP Scale, the most commonly used instrument was the Learning Loss Scale (LLS) developed by Richmond et al. (1987). However, this scale has only two items that measure only the cognitive domain. Using a Likert-type scale, students first are asked to estimate the amount of learning that they gained from a course. The second item asks students to estimate the amount of learning that they would have achieved with the “ideal instructor” (Rovai et al., 2009, p. 9). The scale is scored by subtracting the score for the first item from the second item, creating a learning loss indicator (Richmond et al., 1987). However, because of its brevity, the LLS lacks discriminative value. In other words, the scale measures whether students have learned, but it doesn’t address questions of what, how, and why they learned.

The CAP Scale was developed in three phases. First, the researchers created 80 items that addressed cognitive, affective, and psychomotor learning with between 25 and 28 items per domain (Rovai et al., 2009). In phase one of the study, the scale was administered to 142 higher education students at two universities in Virginia (Rovai et al., 2009). The sample demographics for phase one were not provided. EFA supported a scale with three domains (Rovai et al., 2009). However, some items were cross-loaded. Therefore, in the second phase of the study, the number of questions was reduced to 21 items with seven items per domain, and the scale was administered to 171 new participants (Rovai et al., 2009). The demographics for the phase two sample were not provided. Again, some items were cross-loaded, and the scale was reduced to nine items with three items per domain (Rovai et al., 2009). In phase three, the scale then was administered to a new sample of 221 participants (Rovai et al., 2009). The sample was comprised of 154 females and 67 males. Most students were education majors, and both distance learners ( $n = 157$ ) and traditional classroom students ( $n = 64$ ) were included in the study (Rovai et al., 2009). The ethnicities of the students were as follows: African American ( $n = 36$ ), Asian/Pacific Islander ( $n = 8$ ), Caucasian ( $n = 165$ ), Hispanic ( $n = 6$ ), and other ( $n = 5$ ). Of these, 8 were in the 18-20 age range; 91 were in the 21-30 age range; 63 were in the 31-40 age range; 44 were in the 41-50 age range, and 15 were over 50. The final CAP Scale is comprised of nine 7-point Likert-type items from three domains. Scores for individual items range from 0, *not at all*, to 6, *very much so* (Rovai et al., 2009). Scores on the total scale range from 0 to 54 with scores on each subscale ranging from 0 to 18 (Rovai et al., 2009). Two items are reverse scored. Higher scores correspond with higher levels of perceived learning.

Reliability for the final scale was measured using Cronbach's  $\alpha$ . The internal consistency of the total scale was .79 (Rovai et al., 2009). Test-retest reliability was not reported. Rovai et al. (2009) validated the study using evidence of internal structure by administering it to the student sample described previously. Maximum likelihood CFA was used to assess validity based on internal structure. Most factor loadings were high (greater than .70) or moderately high (greater than .45) (Rovai et al., 2009). The three-factor scale accounted for 67% of the variance (Rovai et al., 2009). The researchers obtained validity evidence based on relations to other variables by comparing the CAP Scale to other measures of perceived learning, including the Affective Learning Scale, the Learning Gain Scale, and the LLS. The inter-correlational matrix of scale items indicated that most items on the scales were correlated. Thus, the researchers concluded that the scale is a valid measure of perceived learning (Rovai et al., 2009). Apart from addressing validity based on test content during the scale development, the researchers did not include validity evidence based on test content, response processes, or consequences of testing. A summary of the three instruments is provided in Table 2.

**Table 2**

*Summary of Instruments*

	CIS	VFLS	CAP Scale
Construct Measured	Motivation	Volition	Perceived learning
Format of Assessment	Survey	Survey	Survey
Reliability	Cronbach's $\alpha = .95$	Cronbach's $\alpha = .825$	Cronbach's $\alpha = .79$
Validity	Reported by Naveed and Bhatti (2017) as possessing acceptable	Reported by Keller et al. (2020) as possessing acceptable	Reported by Rovai et al. (2009) as possessing

	validity evidence based on internal structure using PCA. Other sources of validity not evaluated.	validity evidence based on internal structure using CFA and EFA. Other sources of validity not evaluated.	acceptable validity evidence based on internal structure using CFA and validity evidence based on relations to other variables using an inter-correlational matrix comparing the scale to similar measures of perceived learning. Other sources of validity not evaluated.
Score Range	34-170 (points)	13-65 (points)	0-54 (points)
Subscales	Attention, Relevance, Confidence, Satisfaction	Volition Planning and Volition Control	Cognitive, Affective, Psychomotor

### Procedures

After obtaining Institutional Review Board (IRB) approval, participants were solicited via email on the first day of the semester. A copy of the IRB approval letter is provided in Appendix I. The instructor also posted a general description of the study and an invitation to participate on the course announcements page in D2L. A copy of the recruitment announcement is provided in Appendix J. All students enrolled in the course completed the CIS, VFLS, and CAP Scale. However, an informed consent form with an

opt out option was distributed during the first week of the semester, and only students who agreed to participate in the study were included in the analysis.

Students completed the first set of surveys during the first week of the course to establish a baseline for motivation (attention, relevance, confidence, and satisfaction), volition, and perceived learning, respectively. Participants repeated the three measures during week four of the course and again during week eight of the course. As this was fully-online course, the instruments were administered online through Qualtrics with automated scoring. To ensure fidelity of treatment, the two instructors used the same course content and assessment practices. Identical course content was copied into the D2L course shells for all sections included in the study. Although including students taught by two instructors introduces potential differences in teaching styles and personality traits, the study includes only fully-asynchronous courses with minimal teacher-student interaction, which should minimize these differences. In most cases, the students never interact with the professor in person. Although potential differences in grading practices is a concern, all participants were enrolled in ungraded sections of the course, and both instructors used the same grading criteria. Both instructors also used a grading handbook created by the principal researcher to ensure grading consistency. A copy of the handbook is provided in Appendix K. Additionally, the instructors used the same rubric to assess all major projects. Homework assignments and discussion posts were graded based on completion. Students earned a final grade of A merely by completing assignments with minimal competence, and students were permitted to revise assignments until they met minimum competence. For example, students earned full credit merely for turning in discussion posts that met the minimum word count of 200

words. The rubrics for the three essays and the annotated bibliography are provided in Appendices L-M. These precautions were designed to minimize variance in grading practices.

Because the change in scores over time were compared for each participant, anonymous surveying was not possible. However, students used their student identification numbers rather than their names when completing the surveys. Additionally, the instructors had no access to the survey data until after the study was completed. Single-item and summed scores were computed for each participant for all three scales. Individual scores on the measures at all three time periods were then compared using descriptive statistics. Aggregated cumulative and subscale scores for all participants also were compared.

### **Analysis**

The analysis phase of the study was conducted following the end date for the week eight surveys. Data was imported to Statistical Package for the Social Sciences (SPSS Statistics). Time was the independent variable for this analysis, and scores on the CIS, the VFLS, and the CAP Scale were the dependent variables. Descriptive statistics were used to determine central tendency and variance for total scores and subscale scores at all three time periods. Reliability was assessed using Cronbach's  $\alpha$ . An  $\alpha$  level of greater than .70 was used to determine the reliability of the instruments, as this is the threshold typically applied in social science and education research (Taber, 2018). The appropriateness of parametric analysis was determined through assumption testing. Sphericity was assessed using Mauchly's  $W$ . When sphericity was violated, the

Greenhouse-Geisser correction was applied. The assumption of normality was assessed using a Kolmogorov-Smirnov test.

Each null hypothesis was tested using a within-groups, repeated-measures ANOVA. When the assumption of normality was violated, the results of a Friedman test were compared with the results of the ANOVA. Statistical significance was determined using *p*-values. A significance level of less than .05 was used to determine statistical significance. Partial eta squared ( $\eta^2$ ) was used to determine effect size, which was interpreted using Cohen's (1988) conventions with .01 indicating a small effect, .06 indicating a moderate effect, and .14 indicating a large effect. The results of these analyses were used to determine whether to reject or fail to reject each null hypothesis.

## CHAPTER FOUR: RESULTS

### Introduction

The purpose of this quantitative, survey-based, repeated measures, correlational study was to examine changes in motivation, volition, and perceived learning when ungrading is used in undergraduate research and argumentative writing (RAW) courses. Participants were drawn from seven sections of RAW taught by two instructors during the spring 2023 semester. Three instruments were used in the study. The Course Interest Survey (CIS) was used to measure motivation. The Volition for Learning Scale (VFLS) was used to measure volition, and the CAP Perceived Learning Scale (CAP Scale) was used to measure perceived learning. Data was collected during week one (T<sub>1</sub>), week four (T<sub>2</sub>), and week eight (T<sub>3</sub>) of the semester. A one-way repeated measures analysis of variance (ANOVA) was used for hypothesis testing. When the assumption of normality was untenable, a Friedman test was conducted in addition to the ANOVA, and the results were compared. Because the researchers were interested in changes in motivation, volition, and perceived learning at the subscale level, both total and subscale scores were analyzed for each instrument. This chapter presents the results of the data analysis.

### Results

**Research Question 1: How does student *motivation* (attention, relevance, confidence, and satisfaction) change over time when ungrading is used in online, undergraduate RAW courses?**

#### *CIS Total Scores*

The CIS was used to determine whether motivation changes over time when ungrading is used in undergraduate RAW courses. The instrument is comprised of 34



five-point Likert-type items. Total scores on the measure range from 34 to 170. Cronbach's alpha coefficient (Cronbach's  $\alpha$ ) was used to measure reliability of the CIS for this sample. The reliability of the scale was  $\alpha = .917$  at T<sub>1</sub>,  $\alpha = .906$  at T<sub>2</sub>, and  $\alpha = .938$  at T<sub>3</sub>, which exceeds the threshold of  $\alpha = .70$  suggested in the research literature (e.g., Taber, 2018). Therefore, the reliability of the scale was judged to be acceptable. The descriptive statistics for the measure showed that CIS scores increased slightly during the eight-week study period, as summarized in Table 3. The results of the CIS included a small amount of missing data. One participant failed to complete items 1 and 21 at T<sub>1</sub>. One failed to complete item 10 at T<sub>1</sub>, and one failed to complete item 15 at T<sub>2</sub>. All four missing items were from the attention subscale. The missing data was replaced with the average of the participant's responses on the remaining subscale items.

**Table 3**

*CIS Total Summary of Descriptive Statistics and Reliability Analysis*

	Mean	Std. Deviation	N	Cronbach's Alpha
W1CISTotal	133.56	16.921	57	.917
W4CISTotal	134.58	16.968	57	.906
W8CISTotal	135.93	19.719	57	.938

Assumption testing was used to determine the appropriateness of parametric analysis. The results of the Kolmogorov-Smirnov normality test were  $D(57) = .131, p = .017$  at T<sub>1</sub>,  $D(57) = .130, p = .018$  at T<sub>2</sub>, and  $D(57) = .118, p = .046$  at T<sub>3</sub>. Because the  $p$ -values were below .05, the assumption of normality was rejected. Mauchly's Test of Sphericity (Mauchly's  $W$ ) was used to determine homogeneity of variance. The assumption of sphericity also was rejected,  $\chi^2(2) = 9.474, p = .009$ , so the Greenhouse-Geisser correction was applied when reporting the results of the ANOVA. Although the repeated measures ANOVA is robust over moderate deviations of normality (Morgan et

al., 2019), both the ANOVA and the nonparametric alternative, the Friedman test, were conducted. The results of the ANOVA revealed no significant change in total CIS scores,  $F(1.727, 96.669) = .896, p = .398, \eta^2 = .016$  between T<sub>1</sub> and T<sub>3</sub>. The results showed a small effect with 1.6% of the variance in total CIS scores explained by time. The Friedman test also found no statistically significant changes in total CIS scores from T<sub>1</sub> to T<sub>3</sub>,  $\chi^2(2, n = 57) = 4.283, p = .117$ . Post-hoc tests, with Bonferroni correction for multiple tests, also showed that there were no statistically significant differences in total CIS scores across the three time periods (T<sub>1</sub> to T<sub>2</sub>,  $p = 1.00$ ; T<sub>1</sub> to T<sub>3</sub>,  $p = .132$ ; T<sub>2</sub> to T<sub>3</sub>,  $p = .670$ ). The null hypothesis was retained.

***CIS Attention Subscale***

The attention subscale is comprised of eight Likert-type items. Scores on the subscale range from 8 to 40. Cronbach’s  $\alpha$  was used to measure reliability of the attention subscale. The reliability of the scale was  $\alpha = .764$  at T<sub>1</sub>,  $\alpha = .686$  at T<sub>2</sub>, and  $\alpha = .719$  at T<sub>3</sub>. Although the reliability of the scale at T<sub>2</sub> was suboptimal, the other subscale reliabilities were acceptable. Table 4 presents the descriptive statistics for this subscale. Scores for this subscale decreased slightly from T<sub>1</sub> to T<sub>2</sub> and rebounded from T<sub>2</sub> to T<sub>3</sub>.

**Table 4**

*CIS Attention Subscale Summary of Descriptive Statistics and Reliability Analysis*

	Mean	Std. Deviation	N	Cronbach’s Alpha
W1CISAtt	28.40	5.095	57	.764
W4CISAtt	27.47	5.064	57	.686
W8CISAtt	28.49	4.899	57	.719

Assumption testing was used to determine the appropriateness of parametric analysis. The results of the Kolmogorov-Smirnov normality test for the attention subscale were  $D(57) = .075, p = .200$  at T<sub>1</sub>,  $D(57) = .129, p = .019$  at T<sub>2</sub>, and  $D(57) = .101, p =$

.200 at T<sub>3</sub>. Thus, the assumption of normality was met for T<sub>1</sub> and T<sub>3</sub> and not met for T<sub>2</sub>. Mauchly's W was used to determine homogeneity of variance. The assumption of sphericity also was violated,  $\chi^2(2) = 7.502, p = .023$ . Accordingly, the Greenhouse-Geisser correction was applied when reporting the results of the ANOVA. Because the assumption of normality was violated for T<sub>2</sub> and the assumption of sphericity was rejected for the total subscale, the results of both the repeated measures ANOVA and the Friedman test were considered when answering the research questions. The results of the ANOVA indicated no significant change in the CIS attention subscale scores,  $F(1.774, 99.335) = 1.947, p = .153, \eta^2 = .034$  from T<sub>1</sub> to T<sub>3</sub>. The results showed a small effect with 3.4% of the variance in CIS attention subscale scores explained by time. The Friedman test showed no statistically significant changes in CIS attention subscale scores from T<sub>1</sub> to T<sub>3</sub>,  $\chi^2(2, n = 57) = 2.951, p = .229$ . Post-hoc tests, with Bonferroni correction for multiple tests, showed that there were no statistically significant differences in attention subscale scores across the three time periods (T<sub>1</sub> to T<sub>2</sub>,  $p = .190$ ; T<sub>1</sub> to T<sub>3</sub>,  $p = .851$ ; T<sub>2</sub> to T<sub>3</sub>,  $p = .134$ ). Thus, the null hypothesis was retained.

### ***CIS Relevance Subscale***

The relevance subscale contains nine Likert-type items. Scores on the subscale range from 9 to 45. Cronbach's  $\alpha$  was used to measure the reliability of the subscale with this sample. The reliability of the scale was  $\alpha = .851$  at T<sub>1</sub>,  $\alpha = .804$  at T<sub>2</sub>, and  $\alpha = .888$  at T<sub>3</sub>. The mean scores for the relevance subscale were relatively unchanged from T<sub>1</sub> to T<sub>3</sub>. However, mean scores decreased slightly. Descriptive statistics for the relevance subscale are provided in Table 5 below.

**Table 5***CIS Relevance Subscale Summary of Descriptive Statistics and Reliability Analysis*

	Mean	Std. Deviation	N	Cronbach's Alpha
W1CISRel	37.49	5.504	57	.851
W4CISRel	37.25	5.498	57	.804
W8CISRel	37.18	6.470	57	.888

Assumption testing was performed to determine whether parametric analysis was appropriate. The results of the Kolmogorov-Smirnov normality test for the relevance subscale were  $D(57) = .149, p = .003$  at T<sub>1</sub>,  $D(57) = .158, p = .001$  at T<sub>2</sub>, and  $D(57) = .116, p = .052$  at T<sub>3</sub>. Thus, the subscale scores for T<sub>1</sub> and T<sub>2</sub> were not normally distributed, and the scores for T<sub>3</sub> were normally distributed. Mauchly's W indicated that the assumption of sphericity was not violated,  $\chi^2(2) = 1.645, p = .439$ . Because the assumption of normality was not met at T<sub>1</sub> and T<sub>2</sub>, the Friedman test was performed in addition to the repeated measures ANOVA, and the results were compared. The results of the ANOVA indicated no significant change in CIS relevance subscale scores,  $F(2, 112) = .202, p = .818, \eta^2 = .004$  from T<sub>1</sub> to T<sub>3</sub>. The results showed a negligible effect with .4% of the variance in CIS relevance subscale scores explained by time. The Friedman test showed no statistically significant changes in CIS relevance subscale scores from T<sub>1</sub> to T<sub>3</sub>,  $\chi^2(2, n = 57) = .184, p = .912$ . Post-hoc tests, with Bonferroni correction for multiple tests, also showed that there were no statistically significant differences in scores across the three time periods (T<sub>1</sub> to T<sub>2</sub>,  $p = .963$ ; T<sub>1</sub> to T<sub>3</sub>,  $p = .708$ ; T<sub>2</sub> to T<sub>3</sub>,  $p = .743$ ). Thus, the null hypothesis was retained.

***CIS Confidence Subscale***

The confidence subscale contains eight Likert-type items. Scores on the measure range from 8 to 40. Cronbach's  $\alpha$  was used to measure reliability for the subscale. The

reliability of the scale was  $\alpha = .721$  at T<sub>1</sub>,  $\alpha = .727$  at T<sub>2</sub>, and  $\alpha = .780$  at T<sub>3</sub>. Thus, reliability was judged to be acceptable. The descriptive statistics for the confidence subscale revealed a slight increase in mean scores from T<sub>1</sub> to T<sub>2</sub> and a small decline in mean scores from T<sub>2</sub> to T<sub>3</sub>. The descriptive statistics are summarized in Table 6.

**Table 6**

*CIS Confidence Subscale Summary of Descriptive Statistics and Reliability Analysis*

	Mean	Std. Deviation	N	Cronbach's Alpha
W1CISCon	33.39	4.061	57	.721
W4CISCon	34.61	4.034	57	.727
W8CISCon	34.23	4.888	57	.780

The assumption of normality was tested using the Kolmogorov-Smirnov test. The results were  $D(57) = .157, p = .001$  at T<sub>1</sub>,  $D(57) = .159, p = .001$  at T<sub>2</sub>, and  $D(57) = .168, p < .001$  at T<sub>3</sub>. Because the results showed a normal distribution, a Friedman test was not conducted for this subscale. Mauchly's W revealed that the assumption of sphericity was violated,  $\chi^2(2) = 6.970, p = .031$ . Accordingly, the Greenhouse-Geisser correction was applied when reporting the results of the ANOVA. The results indicated no significant change in CIS confidence subscales scores from T<sub>1</sub> to T<sub>3</sub>,  $F(1.787, 100.087) = .2.772, p = .073, \eta^2 = .047$ . The results showed a medium effect with 7.3% of the variance in CIS confidence subscale scores explained by time. Thus, the null hypothesis was retained.

***CIS Satisfaction Subscale***

Finally, the satisfaction subscale is comprised of nine items. Scores on the instrument range from 9 to 45. Cronbach's  $\alpha$  was used to assess reliability of the subscale. The results were  $\alpha = .713$  at T<sub>1</sub>,  $\alpha = .752$  at T<sub>2</sub>, and  $\alpha = .841$  at T<sub>3</sub>. Reliability was judged to be acceptable for this subscale. The mean scores for the satisfaction

subscale increased slightly across the three time periods. The descriptive statistics for this subscale are presented in Table 7

**Table 7**

*CIS Satisfaction Subscale Summary of Descriptive Statistics and Reliability Analysis*

	Mean	Std. Deviation	N	Cronbach's Alpha
W1CISSat	34.68	4.748	57	.713
W4CISSat	35.25	5.481	57	.752
W8CISSat	36.04	6.153	57	.841

The Kolmogorov-Smirnov test was used to measure normality. The results were  $D(57) = .105, p = .177$  at T<sub>1</sub>,  $D(57) = .148, p = .003$  at T<sub>2</sub>, and  $D(57) = .124, p = .030$  at T<sub>3</sub>. The assumption of normality was confirmed at T<sub>1</sub> and violated at T<sub>2</sub> and T<sub>3</sub>.

Mauchly's W indicated that the assumption of sphericity was not violated,  $\chi^2(2) = 5.204, p = .074$  for the satisfaction subscale. Because the assumption of normality was not met at T<sub>1</sub> and T<sub>2</sub>, the Friedman test was performed in addition to the repeated measures ANOVA, and the results were compared. The results of the ANOVA indicated no significant change in the CIS satisfaction subscale scores,  $F(2, 112) = 2.292, p = .106, \eta^2 = .039$  from T<sub>1</sub> to T<sub>3</sub>. The results showed a small effect with 3.9% of the variance in CIS satisfaction subscale scores explained by time. The Friedman test also showed no statistically significant changes in CIS satisfaction subscale scores from T<sub>1</sub> to T<sub>3</sub>,  $\chi^2(2, n = 57) = 3.972, p = .137$ . Post-hoc tests, with Bonferroni correction for multiple tests, also showed that there were no statistically significant differences in CIS satisfaction subscale scores across the three time periods (T<sub>1</sub> to T<sub>2</sub>,  $p = .454$ ; T<sub>1</sub> to T<sub>3</sub>,  $p = .055$ ; T<sub>2</sub> to T<sub>3</sub>,  $p = .242$ ). The null hypothesis was retained.

**Research Question 2: How does student *volition* (volition planning and volition control) change over time when ungrading is used in online, undergraduate RAW courses?**

***VFLS Total Scores***

The VFLS was used to answer the second research question regarding changes in volition from T<sub>1</sub> to T<sub>3</sub>. The scale contains 13 five-point Likert-type items. Scores on the VFLS range from 5 to 65. Cronbach’s  $\alpha$  was used to measure reliability of the VFLS for this sample. The reliability of the scale was  $\alpha = .888$  at T<sub>1</sub>,  $\alpha = .876$  at T<sub>2</sub>, and  $\alpha = .902$  at T<sub>3</sub>. The reliability of the scale was judged to be acceptable. Total mean scores decreased slightly from T<sub>1</sub> to T<sub>2</sub> and increased slightly from T<sub>2</sub> to T<sub>3</sub>. The means for the total scores are presented in Table 8.

**Table 8**

*VFLS Total Summary of Descriptive Statistics and Reliability Analysis*

	Mean	Std. Deviation	N	Cronbach’s Alpha
W1VFLSTotal	55.33	7.395	57	.888
W4VFLSTotal	52.77	7.481	57	.876
W8VFLSTotal	53.95	8.164	57	.902

Assumption testing was used to determine the appropriateness of parametric analysis. The Kolmogorov-Smirnov test was used to judge normality. The results were  $D(57) = .096, p = .200$  at T<sub>1</sub>,  $D(57) = .109, p = .092$  at T<sub>2</sub>, and  $D(57) = .103, p = .200$  at T<sub>3</sub>. Because each of these values exceeds the .05 threshold, the scores were found to be normally distributed. Mauchly’s W indicated that the assumption of sphericity was not violated for the total VFLS scores,  $\chi^2(2) = .258, p = .879$ . Given the results of assumption testing, a repeated measures ANOVA was used to test changes in mean scores for the subscale. The results showed a statistically significant change in volition from T<sub>1</sub> to T<sub>3</sub>,

$F(2, 112) = 4.859, p = .009, \eta^2 = .080$ . The results showed a medium effect with 8.0% of the variance in total VFLS scores explained by time. The null hypothesis was rejected.

***VFLS VP Subscale***

The VP subscale is comprised of five Likert-type items. Scores on the scale range from 5 to 25. Cronbach’s  $\alpha$  was used to measure the reliability of the VP subscale. The results were  $\alpha = .782$  at T<sub>1</sub>,  $\alpha = .823$  at T<sub>2</sub>, and  $\alpha = .877$  at T<sub>3</sub>. The reliability of the subscale was acceptable. Scores on the VP subscale decreased slightly from T<sub>1</sub> to T<sub>3</sub>. The descriptive statistics for the VP subscale are presented in Table 9.

**Table 9**

*VFLS Volition Planning Summary of Descriptive Statistics and Reliability Analysis*

	Mean	Std. Deviation	N	Cronbach’s Alpha
W1VFLSVP	22.05	2.856	57	.782
W4VFLSVP	21.30	2.952	57	.823
W8VFLSVP	21.75	3.356	57	.877

The Kolmogorov-Smirnov test was used to measure the assumption of normality for the VP subscale. The results of the test were  $D(57) = .173, p < .001$  at T<sub>1</sub>,  $D(57) = .139, p = .008$  at T<sub>2</sub>, and  $D(57) = .189, p < .001$  at T<sub>3</sub>, indicating that the data was not normally distributed. Therefore, the results of the Friedman test were reported in addition to the ANOVA, and the results were compared. Mauchly’s W indicated that the assumption of sphericity was not violated,  $\chi^2(2) = .982, p = .612$ . Therefore, a repeated measures ANOVA was used to test the null hypothesis. The results of the ANOVA showed that there was no significant change in VP from T<sub>1</sub> to T<sub>3</sub>,  $F(2, 112) = 2.386, p = .097, \eta^2 = .041$ . The results showed a small effect with 4.1% of the variance in VP subscale scores explained by time. The Friedman test also showed no statistically significant changes in VP subscale scores from T<sub>1</sub> to T<sub>3</sub>,  $\chi^2(2, n = 57) = 3.801, p = .149$ .



Post-hoc tests, with Bonferroni correction for multiple tests, also showed that there were no statistically significant differences in VP subscale scores across the three time periods,  $T_1$  to  $T_2$ ,  $p = .101$ ;  $T_1$  to  $T_3$ ,  $p = .640$ ;  $T_2$  to  $T_3$ ,  $p = .242$ . The null hypothesis was retained.

***VFLS VC Subscale***

The VC subscale contains eight Likert-type items. Scores on the measure range from 8 to 40. Cronbach’s  $\alpha$  was used to measure the reliability of the VC subscale. The results were  $\alpha = .823$  at  $T_1$ ,  $\alpha = .794$  at  $T_2$ , and  $\alpha = .832$  at  $T_3$ . The reliability of the scale was acceptable. The means scores for the VC subscale revealed a slight decrease in mean scores from  $T_1$  to  $T_3$ . Table 10 presents the descriptive statistics for this subscale.

**Table 10**

*VFLS Volition Control Summary of Descriptive Statistics and Reliability Analysis*

	Mean	Std. Deviation	N	Cronbach’s Alpha
W1VFLSVC	33.28	4.924	57	.823
W4VFLSVC	31.47	5.040	57	.794
W8VFLSVC	32.19	5.347	57	.832

The Kolmogorov-Smirnov was used to assess the assumption of normality. The results were  $D(57) = .134$ ,  $p = .012$  at  $T_1$ ,  $D(57) = .092$ ,  $p = .200$  at  $T_2$ , and  $D(57) = .106$ ,  $p = .171$  at  $T_3$ . The assumption of normality was violated at  $T_1$  and upheld at  $T_2$  and  $T_3$ . Mauchly’s  $W$  indicated that the assumption of sphericity was not violated for the VC subscale,  $\chi^2(2) = 1.476$ ,  $p = .478$ . Because the data was not normally distributed at  $T_1$ , the Friedman test was reported in addition to the ANOVA, and the results were compared. The results of the ANOVA revealed a statistically significant change in VC subscale scores from  $T_1$  to  $T_3$ ,  $F(2, 112) = 5.152$ ,  $p = .007$ ,  $\eta^2 = .084$ . The results showed a medium effect with 8.4% of the variance in VC subscale scores explained by time. The Friedman test also showed a statistically significant change in VC subscale scores from

T<sub>1</sub> to T<sub>3</sub>,  $\chi^2(2, n = 57) = 10.207, p = .006$ . Post-hoc tests, with Bonferroni correction for multiple tests, also showed that there was a statistically significant difference in VC scores across the three time periods (T<sub>1</sub> to T<sub>2</sub>,  $p = .004$ ; T<sub>1</sub> to T<sub>3</sub>,  $p = .031$ ; T<sub>2</sub> to T<sub>3</sub>,  $p = .454$ ). The null hypothesis was rejected.

**Research Question 3: How does student *perceived learning* (cognitive, affective, psychomotor) change over time when ungrading is used in online, undergraduate RAW courses?**

***CAP Scale Total Scores***

The CAP Scale was used to measure changes in perceived learning across the three time periods. The total scale is comprised of nine, six-point Likert-type items. Scores on the measure range from 0 to 54. The reliability of the CAP Scale was measured by Cronbach’s  $\alpha$ . The results showed that the scale had acceptable reliability with this sample,  $\alpha = .878$  at T<sub>1</sub>,  $\alpha = .793$  at T<sub>2</sub>, and  $\alpha = .794$  at T<sub>3</sub>. The results of the CAP Scale included a small amount of missing data. One participant failed to complete item 1 at T<sub>1</sub>. The missing data was replaced with the average of the participant’s scores on the remaining cognitive subscale items. The mean total scores decreased from T<sub>1</sub> to T<sub>3</sub>. The descriptive statistics for the total score are presented in Table 11.

**Table 11**

*CAP Scale Total Summary of Descriptive Statistics and Reliability Analysis*

	Mean	Std. Deviation	N	Cronbach’s Alpha
W1CAPTotal	43.93	5.685	57	.878
W4CAPTotal	40.35	7.042	57	.793
W8CAPTotal	40.23	7.419	57	.794

The assumption of normality was assessed using the Kolmogorov-Smirnov test. The results were  $D(57) = .127, p = .023$  at T<sub>1</sub>,  $D(57) = .107, p = .164$  at T<sub>2</sub>, and  $D(57) =$

.101,  $p = .200$  at  $T_3$ . Mauchly's  $W$  showed that the assumption of sphericity for the total scale was violated,  $\chi^2(2) = 13.553$ ,  $p = .001$ , the Greenhouse-Geisser correction was applied. Because the assumption of normality was violated for the  $T_1$  scores, a Friedman test was performed in addition to the repeated measures ANOVA, and the results were compared. The results of the ANOVA revealed a statistically significant change in perceived learning from  $T_1$  to  $T_3$ ,  $F(1.641, 91.923) = 13.513$ ,  $p < .001$ ,  $\eta^2 = .194$ . The results showed a large effect with 19.4% the variance in total CAP Scale scores explained by time. The Friedman test also showed a statistically significant change in total CAP Scale scores from  $T_1$  to  $T_3$ ,  $\chi^2(2, n = 57) = 17.689$ ,  $p < .001$ . Post-hoc tests, with Bonferroni correction for multiple tests, also showed that there was a statistically significant difference in total CAP Scale scores across the three time periods,  $T_1$  to  $T_2$ ,  $p < .001$ ;  $T_1$  to  $T_3$ ,  $p < .001$ ;  $T_2$  to  $T_3$ ,  $p = 1.000$ . The null hypothesis was rejected.

### ***CAP Scale Cognitive Subscale***

The cognitive subscale is comprised of three Likert-type items. Scores on the scale range from 0 to 18. Cronbach's  $\alpha$  was used to assess the reliability of the cognitive subscale. The results were  $\alpha = .691$  at  $T_1$ ,  $\alpha = .363$  at  $T_2$ , and  $\alpha = .305$  at  $T_3$ . The reliabilities of the subscale items for  $T_2$  and  $T_3$  were quite low. Therefore, an inter-item correlation matrix was used to explore the relationships among the subscale items for  $T_2$  and  $T_3$ . Table 12 presents the inter-item correlations. As shown in the table, all three pairs decreased from  $T_1$  to  $T_3$ . Although the low reliability indicates a potential for measurement error in the analysis, the analysis was continued for the sake of completeness of reporting.

**Table 12***CAP Scale Cognitive Subscale Inter-Item Correlation Matrix*

Item Pairs	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>
CAP1 with CAP2	.404	.170	.055
CAP1 with CAP5	.336	.301	.202
CAP2 with CAP5	.536	.119	.193

The descriptive statistics for the cognitive subscale are presented in Table 13. The results indicated a slight increase from T<sub>1</sub> to T<sub>2</sub> and a slight decrease from T<sub>2</sub> to T<sub>3</sub>.

**Table 13***CAP Scale Cognitive Subscale Summary of Descriptive Statistics and Reliability Analysis*

	Mean	Std. Deviation	N	Cronbach's Alpha
W1CAPCog	11.26	1.158	57	.691
W4CAPCog	13.16	2.463	57	.363
W8CAPCog	12.68	2.237	57	.305

The Kolmogorov-Smirnov normality test was used to evaluate the assumption of normality. The results were  $D(57) = .125, p = .027$  at T<sub>1</sub>,  $D(57) = .124, p = .030$  at T<sub>2</sub>,  $D(57) = .158, p = .001$  at T<sub>3</sub>. Thus, the assumption of normality was violated for this subscale. Mauchly's W determined that the assumption of sphericity was not violated for the cognitive subscale,  $\chi^2(2) = 2.367, p = .306$ . Because the assumption of normality was violated, a Friedman test was performed in addition to the ANOVA, and the results were compared. The ANOVA indicated a statistically significant change in score across the three time periods,  $F(2, 112) = 27.480, p < .001, \eta^2 = .147$ . The results showed a large effect with 14.7% of the variance in CAP Scale cognitive subscale scores explained by time. The Friedman test also showed a statistically significant change in CAP Scale cognitive subscale scores from T<sub>1</sub> to T<sub>3</sub>,  $\chi^2(2, n = 57) = 13.519, p = .001$ . Post-hoc tests, with Bonferroni correction for multiple tests, also showed that there were statistically

significant differences in subscale scores across the three time periods ( $T_1$  to  $T_2$ ,  $p = .001$ ;  $T_1$  to  $T_3$ ,  $p = .015$ ;  $T_2$  to  $T_3$ ,  $p = .454$ ). The null hypothesis was rejected.

### ***CAP Scale Affective Subscale***

The affective subscale contains three Likert-type items. Scores on the scale range from 0 to 18. Cronbach's  $\alpha$  was used to determine the reliability of the subscale. The results were .749 at  $T_1$ , .778 at  $T_2$ , and .720 at  $T_3$ . The reliability of the subscale was acceptable. For the affective subscale, scores decreased from  $T_1$  to  $T_2$  and increased from  $T_2$  to  $T_3$ . Means scores for this subscale are provided in Table 14.

**Table 14**

#### *CAP Scale Affective Subscale Summary of Descriptive Statistics and Reliability Analysis*

	Mean	Std. Deviation	N	Cronbach's Alpha
W1CAPAff	14.14	2.364	57	.749
W4CAPAff	13.12	3.235	57	.778
W8CAPAff	13.60	3.081	57	.720

The Kolmogorov-Smirnov normality test was used to evaluate the assumption of normality. The results were  $D(57) = .098$ ,  $p = .200$  at  $T_1$ ,  $D(57) = .152$ ,  $p = .002$  at  $T_2$ , and  $D(57) = .119$ ,  $p = .044$  at  $T_3$ . The assumption of normality was accepted at  $T_1$  and rejected at  $T_2$  and  $T_3$ . Mauchly's  $W$  indicated that the assumption of sphericity was not violated,  $\chi^2(2) = 5.106$ ,  $p = .078$ . Because the assumption of normality was rejected at  $T_2$  and  $T_3$ , a Friedman test was performed, and the results were compared. The ANOVA revealed a statistically significant change in affective perceived learning from  $T_1$  to  $T_3$ ,  $F(2, 112) = 3.579$ ,  $p = .031$ ,  $\eta^2 = .060$ . The results showed a medium effect with 6.0% of the variance in CAP Scale affective subscale scores explained by time. The Friedman test showed no statistically significant changes in CAP Scale affective subscale scores from  $T_1$  to  $T_3$ ,  $\chi^2(2, n = 57) = 5.225$ ,  $p = .073$ . Post-hoc tests, with Bonferroni correction for

multiple tests, also showed that there were no statistically significant differences in CAP cognitive subscale scores across the three time periods (T<sub>1</sub> to T<sub>2</sub>,  $p = .044$ ; T<sub>1</sub> to T<sub>3</sub>,  $p = .606$ ; T<sub>2</sub> to T<sub>3</sub>,  $p = .134$ ). The results of the ANOVA and the Friedman's test differed for this subscale. The null hypothesis was retained, as it was the more conservative option.

### ***CAP Scale Psychomotor Subscale***

The psychomotor subscale is comprised of three Likert-type items. Scores on the subscale range from 0 to 18. Cronbach's  $\alpha$  coefficient was used to determine reliability for the psychomotor subscale. The results were  $\alpha = .869$  at T<sub>1</sub>,  $\alpha = .733$  at T<sub>2</sub>, and  $\alpha = .730$  at T<sub>3</sub>. The reliability of the scale was acceptable. For the psychomotor subscale, mean scores increased from T<sub>1</sub> to T<sub>2</sub> and decreased from T<sub>2</sub> to T<sub>3</sub>. The descriptive statistics for this subscale are presented in Table 15.

**Table 15**

#### *CAP Scale Psychomotor Subscale Summary of Descriptive Statistics and Reliability*

##### *Analysis*

	Mean	Std. Deviation	N	Cronbach's Alpha
W1CAPPsy	15.40	2.052	57	.869
W4CAPPsy	14.07	3.133	57	.733
W8CAPPsy	13.42	3.474	57	.730

The Kolmogorov-Smirnov normality test was used to examine the assumption of normality of the subscale. The results were  $D(57) = .176$ ,  $p < .001$  at T<sub>1</sub>,  $D(57) = .196$ ,  $p < .001$  at T<sub>2</sub>, and  $D(57) = .149$ ,  $p = .003$  at T<sub>3</sub>. The assumption of normality was rejected. Mauchly's W indicated that the assumption of sphericity was not violated for the psychomotor subscale,  $\chi^2(2) = 4.606$ ,  $p = .100$ . Because the assumption of normality was rejected, a Friedman test was completed as well, and the results were compared. The results of the ANOVA showed that there was a statistically significant change in scores

from T<sub>1</sub> to T<sub>3</sub>,  $F(2, 112) = 12.617, p < .001, \eta^2 = .184$ . The results showed a large effect with 18.4% of the variance in CAP psychomotor subscale scores explained by time. The results of the Friedman test also indicated that there was a statistically significant change in psychomotor perceived learning from T<sub>1</sub> to T<sub>3</sub>,  $\chi^2(2, n = 57) = 19.198, p < .001$ . Post-hoc tests, with Bonferroni correction for multiple tests, also showed that there were no statistically significant differences in subscale scores across the three time periods (T<sub>1</sub> to T<sub>2</sub>,  $p = .004$ ; T<sub>1</sub> to T<sub>3</sub>,  $p < .001$ ; T<sub>2</sub> to T<sub>3</sub>,  $p = .374$ ). The null hypothesis was rejected.

### **Summary**

This study examined changes in student motivation, volition, and perceived learning when ungrading is used in undergraduate RAW courses. Fifty-seven students participated in the study. Motivation was measured using the CIS. Volition was measured using the VFSL, and perceived learning was measured using the CAP Scale. Reliability for each scale and subscale was measured using Cronbach's  $\alpha$ . The results indicated acceptable reliability for most scales. However, the reliability of the cognitive subscale from the CAP Scale had suboptimal reliability.

Assumption testing for each scale and subscale was performed prior to hypothesis testing. The assumption of normality was assessed using the Kolmogorov-Smirnov test, and the assumption of sphericity was assessed using the Mauchly's  $W$ . When the assumption of sphericity was violated, the Greenhouse-Geisser correction was applied. When the assumption of normality was violated, the results of the Friedman test were reported in tandem with the ANOVA results to provide a more holistic view of the data. A summary of the assumption testing is provided in Table 16.

**Table 16***Summary of Normality and Sphericity Assumption Testing*

Scale/Subscale	Normality Assumption			Sphericity Assumption
	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>1</sub> – T <sub>3</sub>
CIS Total	Not met	Not met	Not met	Not met
Attention	Met	Not met	Met	Not met
Relevance	Not met	Not met	Met	Met
Confidence	Not Met	Not Met	Not Met	Not met
Satisfaction	Met	Not met	Not met	Met
VFLS Total	Met	Met	Met	Met
Volition Planning	Not met	Not met	Not met	Met
Volition Control	Not met	Met	Met	Met
CAP Scale Total	Not met	Met	Met	Not met
Cognitive	Not met	Not met	Not met	Met
Affective	Met	Not met	Not met	Met
Psychomotor	Not met	Not met	Not met	Met

The null hypotheses were examined using a repeated measures ANOVA and the nonparametric equivalent of the ANOVA, the Friedman test. The hypothesis testing for the first research question showed no statistically significant change in motivation for either the total CIS scale or its subscales. Therefore, the null hypotheses for this research question were retained. Regarding the second research question, the results showed a statistically significant change in total VLFS scores from T<sub>1</sub> to T<sub>3</sub>. For the subscales, the results showed a statistically significant change for VP but no statistically significant change for VC. Thus, the null hypotheses were rejected for both the total VLFS scale and the VC subscale and retained for the VP subscale. For the third research question, the results showed a statistically significant change for the total CAP Scale and the cognitive



and psychomotor subscales and no statistically significant change for the affective subscale. The null hypotheses were rejected for both the total CAP Scale and the cognitive and psychomotor subscales, and the null hypothesis was retained for the affective subscale. A summary of the hypothesis testing results is provided in Table 17.

**Table 17**

*Summary of ANOVA and Friedman Test Results*

	ANOVA ( $p < .05$ )	Friedman ( $p < .05$ )
Scale/Subscale	T <sub>1</sub> – T <sub>3</sub>	T <sub>1</sub> – T <sub>3</sub>
CIS Total	Not significant, small effect	Not significant
Attention	Not significant, small effect	Not significant
Relevance	Not significant, negligible effect	Not significant
Confidence	Not significant, small effect	N/A
Satisfaction	Not significant, small effect	Not significant
VFLS Total	Significant, medium effect	N/A
Volition Planning	Not significant, small effect	Not significant
Volition Control	Significant, medium effect	Significant
CAP Scale Total	Significant, large effect	Significant
Cognitive	Significant, large effect	Significant
Affective	Significant, medium effect	Not significant
Psychomotor	Significant, large effect	Significant

*Note:* Not significant indicates that the  $p$ -value was less than .05.

## CHAPTER FIVE: DISCUSSION

### Introduction

The purpose of this quantitative, survey-based, repeated measures, correlational study was to examine changes in motivation (attention, relevance, confidence, and satisfaction), volition (volition planning and volition control) and perceived learning (cognitive, affective, and psychomotor) when ungrading is used in undergraduate, online research and argumentative writing (RAW) courses. Ungrading is an umbrella term for a variety of assessment practices. The underlying tenet of all forms of ungrading is that instructors should provide extensive formative feedback rather than summative grades. Ungrading is intended to engender a better learning experience by increasing affective characteristics, such as motivation, volition, and perceived learning. The ARCS-V model (Keller, 1979; Keller, 1983; Keller, 1987; Keller, 2008a; Keller, 2008b; Keller, 2010; Keller et al., 2020) served as the theoretical framework for the study. The model posits that learning increases when motivation and volition are present (Keller, 2010; Li & Keller, 2018). Keller et al. (2020) defined motivation as “having an objective and developing plans to achieve this objective,” whereas volition involves acting and ultimately achieving the objective (p. 161). Thus, volition follows motivation because one acts after becoming motivated (Keller et al., 2020). As a strategy for increasing motivation, volition, and perceived learning, ungrading aligns with this model. To understand the potential impact of ungrading, motivation was measured using the Course Interest Survey (CIS) (RQ1). Both total scores and subscale scores (attention, relevance, confidence, satisfaction) were reported. Volition was measured using the Volition for Learning Scale (VFLS) (RQ2). This scale contains two subscales (volition planning and

volition control). Finally, perceived learning was measured using the CAP Perceived Learning Scale (CAP Scale), which contains three subscales (cognitive, affective, and psychomotor) (RQ3).

Participants for this study were derived from seven sections of RAW taught by two non-tenure-track instructors during the spring 2023 semester. A repeated measures design was used for this study because it allowed the researchers to assess changes in motivation, volition, and perceived learning over time. Participants completed the CIS, the VFLS, and the CAP Scale during week one (T<sub>1</sub>), week four (T<sub>2</sub>), and week 8 (T<sub>3</sub>) of the course. The final sample included 57 students. The intervention for the study was a labor-based grading system in which final grades are awarded based on labor rather than achievement. In alignment with this ungrading strategy, students received a grade of *accept* or *revise* for individual assignments. Students who completed all assignments at the *accept* level earned a letter grade of A in the course. This chapter presents a discussion and interpretation of the principal findings.

## Discussion

**Research Question 1: How does student *motivation* (attention, relevance, confidence, and satisfaction) change over time when ungrading is used in online, undergraduate RAW courses?**

Research question one examined changes in motivation over time, as measured by the CIS, when ungrading is employed in undergraduate, online RAW courses. The results of the study indicated that there was no statistically significant change in motivation over time when ungrading was used. This finding was true for both total CIS scores (RQ1) and subscale scores (RQ1.1 - 1.4). These results were unanticipated, given the extensive body

of literature suggesting that ungrading increases motivation. With that said, it is important to note that motivation scores did not decline from T<sub>1</sub> to T<sub>3</sub>; they remained constant (RQ1). Thus, total motivation neither increased nor decreased. Also, a within-subjects design was used, so changes in motivation across all students were not measured. Instead, students' responses were paired with their prior responses. Several factors may explain these anomalous results.

First, research suggests that academic motivation tends to be static, especially over short time periods (An & Li, 2021; Foong et al., 2022; Sulz et al., 2020). Although most studies into academic motivation have found that motivation changes in response to motivational teaching strategies, there is some precedent for this finding. For example, Foong et al. (2022) studied motivation over time using a sample of 46 undergraduate first-year engineering students. As in this study, they found that student motivation did not change during a six-month period (Foong et al., 2022). They noted that intrinsic motivation was especially resistant to change. In another study of medical students, An and Li (2021) found that motivation was invariant over time. Collectively, these findings suggest that academic motivation can be difficult to change. A related possibility is that eight weeks might not be enough time to produce observable changes in motivation, which aligns with some prior research. For example, Foong et al. (2022) followed students for one semester. In a similar study concerning motivation among tenth graders enrolled in a health class, Sulz et al. (2020) found no change in motivation after four months. In contrast to cognitive learning outcomes, it is possible that longer time periods are necessary to produce statistically significant changes.

Academic motivation also could be subject to a ceiling effect, which may have attenuated the positive impacts of ungrading. The ceiling effect occurs when scores on an instrument are already high at the beginning of the study, which limits the amount of change that can occur (Staus et al., 2021). Many studies have observed this effect within learning contexts. For example, Status et al. (2021) observed a ceiling effect in their study of an afterschool STEM camp, and Rifkin (2005) found that students in a summer language-learning classroom experienced a ceiling effect. As it relates to socio-emotional learning language outcomes, Douglas et al., (2020) argued that assessing motivation using self-report measures often leads to a ceiling effect. Because scores on the CIS were already rather high at T<sub>1</sub> ( $M = 133.56$  out of 170 possible points), there may have been limited range for improvement, and the instrument may not have been sensitive to small fluctuations in motivation.

Next, the findings of previous studies into academic motivation may not extend to assessment. The present study was unique in that it examined the motivational effects of an assessment strategy rather than an instructional strategy. The literature review revealed no studies that have examined ungrading in the context of the ARCS-V model. Common interventions represented in the ARCS-V literature include project-based learning (Alhassan, 2014; Chao et al., 2019; Jeon et al., 2014), educational games (Hao, & Lee, 2019; Karakiş et al., 2016; Proske et al., 2017; Sdravopoulou et al., 2021; Wah, 2015), and podcasts (Moura & Carvalho, 2012). In each of these studies, the ARCS-V model was focused on the learning intervention and learning resources as opposed to an assessment strategy. In addition to using different interventions, other studies into

motivation have targeted different constructs. Thus, it is possible that assessment strategies have little effect on overall motivation (RQ1).

It also is possible that students viewed ungrading as traditional grading by a different name, which tempered the motivational benefits of ungrading (McMorran & Ragupathi, 2020). Students may not have perceived the grading system as distinctive or different from traditional grades. This could undermine their perceptions of the benefits of ungrading, as some research suggests that students often do not understand what ungrading entails or how it differs from traditional grading, and some researchers and scholars have argued that ungrading is merely traditional grading with different terminology. For example, McMorran and Ragupathi (2020) found that some students in their study viewed ungrading as delaying the inevitability of a final grade. Thus, the insignificant results might stem from students' perceptions that ungrading is not meaningfully distinct from traditional grading.

**Research Question 2: How does student *volition* (volition planning and volition control) change over time when ungrading is used in online, undergraduate RAW courses?**

### ***Implications for Ungrading***

Research question two examined changes in volition over time when ungrading is used in undergraduate, online RAW courses. In contrast to the motivation learning outcome (RQ1), the results of the study indicated that there was a statistically significant decline in total volition (RQ2) when ungrading was employed. This conclusion was true for total VFLS scores (RQ2) and for scores on the volition control (VC) subscale (RQ 2.2). There was no statistically significant change for the volition planning (VP) subscale

(RQ2.1). For the VFLS total scale and the VC subscale, scores decreased from T<sub>1</sub> to T<sub>2</sub> and modestly rebounded from T<sub>2</sub> to T<sub>3</sub> (RQ2.2). However, scores at T<sub>3</sub> were still lower than scores at T<sub>1</sub>, suggesting that volition control decreased from T<sub>1</sub> to T<sub>3</sub> and that this change was statistically significant (RQ2.2). These results were unexpected given a large body of ungrading research, which suggests that ungrading increases total volition. Indeed, prior studies have determined that ungrading benefits students by increasing agency, autonomy, and self-regulation, which are critical components of both VP and VC (Brubaker, 2010; Gorichanaz, 2022; Lindemann & Harbke, 2011; Pulfrey et al., 2011). By contrast, the results of this study indicate that students did not experience enhanced total volition when ungrading was employed (RQ2).

There may be multiple interpretations for this finding. First, the timing of both the surveys and formative feedback may have contributed to these results. Koenka et al. (2019) found that the effects of feedback vary depending on when it is provided. In this study, the surveys were timed to coincide with certain milestones. Participants completed the first set of surveys during the first week of classes before the first major essay of the semester had been assigned. They completed the second set of surveys immediately prior to the due date for the first major essay and completed the third set of surveys after receiving formative feedback for the first essay. It is possible that the mental state of anticipating feedback differs from the experience receiving feedback, which may have reduced VC, as some students may have been discouraged upon receiving feedback at T<sub>2</sub> (RQ2.2). This possibility is especially relevant given that some students were required to revise and resubmit their essays while simultaneously completing other assignments. This aligns with some prior research. For example, Chamberlin et al. (2018) found that interest

declines following the assessment task, especially when the academic subject is not intrinsically interesting to students. Because VP is predicated upon interest in the task (RQ2.1), declining interest following an assessment may decrease VC (RQ2.2). Thus, the assessment stage may have impacted total volition.

Another issue is that students might have perceived ungrading as vague or subjective, which led to increased anxiety and perceived loss of control, which may have decreased both VP (RQ2.1) and VC (RQ2.2). In the present study, most students were unfamiliar with ungrading before enrolling in this course, and students may have become reticent when presented with an unfamiliar teaching strategy. There is some support for this phenomenon. For example, Mallette and Hawks (2020) found that some students felt that grading contracts were unclear and preferred the traditional grading system. Similarly, Reardon and Guardado-Menjivar (2020) found that some students disliked ungrading because they felt unable to assess their performance, preferring grades so that they could track their progress. Finally, Inman and Powell (2018) found that grades are profoundly meaningful for students and that they inform certain aspects of students' identities. Thus, grading preferences may have resulted in resistance to ungrading, which led to decreased total volition to participate in it.

It is notable that participants showed a statistically significant decrease in VC (RQ2.2) but not VP (RQ2.1). As defined in the ARCS-V model, VP (RQ2.1) involves becoming attracted to a goal and committing to action, while VC (RQ2.2) involves taking action to complete the goal (Keller, 2008a; Keller, 2008b; Keller, 2010). Thus, VP (RQ2.1) is primarily a mental process, and VC (RQ2.2) is a demonstrable action. In the ARCS-V model, VP (RQ2.1) is a precursor to VC (RQ2.2). (Keller, 2008a; Keller,



2008b; Keller, 2010). Both constructs are closely related to autonomy, agency, and perceived control, which are antecedents of volition, given that students must be mentally prepared to act (VP) before acting (VC). Ostensibly, encouraging VP (RQ2.1) should correlate with increased VC (RQ2.2). The findings of this study are significant because they indicate that VC (RQ2.2) constructs may have decreased over time in the absence of a concordant decrease in planning and preparation (RQ2.1). These results align with some prior research, which shows that ungrading might decrease certain aspects of total volition. As Hattie and Temperley (2007) have argued, students conduct a cost benefit analysis when determining how much volition to exert. When students perceive that the affective costs outweigh the potential benefits, they experience avoidance and procrastination, which reduces VC (Hattie & Timperley, 2007). Thus, it is possible that students perceived ungrading as inhibiting.

Next, it is possible that ungrading resulted in complacency with this sample of students, which reduced volition. There is some precedent for this assertion. For example, Merva (2003) found that students exerted less effort and exhibited less preparation when pass-fail ungrading was used, arguing that some students view ungraded courses as an “academic vacation” (p. 155). Michaelides and Kirshner (2005) also found that students exert more effort in graded courses, as did McMorran & Ragupathi (2020). Similarly, McMorran et al. (2017) found that it can be difficult to convince students that learning is intrinsically meaningful in the absence of grades. Thus, although there is a substantial body of research that supports the efficacy of ungrading, some researchers have found contradictory results. Research suggests that grades are inherently volitional for some students, and it is possible that the absence of grades reduced volition.

It is important to acknowledge that no grading system is equally volitional for all students (Cowan, 2020). Although much of the literature in the field describes the benefits of ungrading, a substantial minority of studies have found that students are resistant to ungrading. This issue has been addressed in research literature. For example, Nystrom (2018) found that students who expected to earn high grades were likely to prefer traditional grades. Ferguson (2016) also found that high performing students preferred traditional grades, while lower performing students preferred ungrading. Conversely, Barber (2021) found that most students, regardless of ability, preferred traditional grades. Some students also felt that ungrading was unfair. As Spidell and Thelin (2006) found, high performers resented the fact that low performers were permitted to revise assignments. Similarly, McMorran and Ragupathi (2020) found that ungrading both increased and decreased stress, depending on students' abilities. Collectively, studies show that student grading preferences are diverse. It is possible that this sample of students did not perceive ungrading as volitional.

### ***Ungrading in Online Learning for RAW Students***

The results also are noteworthy because they extend the research on ungrading to include the online learning setting. Another potential reason for these unanticipated findings is that the online learning environment may have tempered the benefits of ungrading, contributing to decreased volition. There is limited research into the volitional impacts of ungrading in online courses, as most research into ungrading has taken place in traditional classroom settings (e.g., Brubaker, 2010; Bullock et al., 2022; Chamberlin et al., 2018; Inman & Powell, 2018; Landrum & Dietz, 2006; Lindemann & Harbke, 2011; Litterio, 2016; Litterio, 2018; Michaelides & Kirshner, 2005; Pulfrey et al., 2011;

Roberts & Dorstyn, 2017; Rohe et al., 2006; Seligman et al., 2021; Spidell & Thelin, 2006; Ward, 2021a; Ward, 2021b; White & Fantone, 2010). Promoting total volition (RQ2) in online courses is especially challenging, as some students are unprepared for the time management, self-discipline, and self-efficacy required for successful online learning (Firat et al., 2018; Hobson & Puruhito, 2018; Park & Yun, 2017; Wang et al., 2008). This problem is especially evident in first-year, general education courses, such as RAW. It is possible that using ungrading in online courses is less effective for increasing total volition (RQ2).

A related issue is that online learning can be isolating for some students, which may have decreased total volition, as lack of social presence is a potential barrier to volition in online courses (Elkins, 2016; Horstmanshof & Brownie, 2013; Neal, 2011; Sapp & Simon, 2005). Researchers have noted that ungrading redistributes power and control and encourages an equitable collaborative relationship between instructors and their students, increasing student volition by increasing agency, autonomy, and self-efficacy (Brubaker, 2010; Cowan, 2020; Ferguson, 2013; Guberman, 2021; Hiller & Hietapelto, 2001; Litterio, 2016; Stuckey et al., 2020; Tannock, 2017). Despite these benefits, there are inherent, and largely unavoidable, physical barriers in online education. Sapp and Simon (2005) reported that their online students felt that the online environment did not facilitate students' relationships with their professor (53%) or their peers (79%). Although students in on-campus courses have varying levels of interaction with the instructor, these courses typically offer students an opportunity to interact with both their instructor and their peers multiple times a week, allowing students to ask questions and receive immediate feedback from multiple sources. In-person learning also

allows for real-time communication. Students may ask follow-up questions and request clarification, without the delay that accompanies text-based communication. Conversely, students in online classes may feel detached from the instructor and their peers because there is little personal interaction, even when periodic videoconferencing is incorporated. The present study indicates that the potential volitional benefits of ungrading might be tempered by the online environment because students might have felt isolated due to a lack of social presence (RQ2).

It also is possible that lack of course interest impeded the volitional benefits of ungrading. The study took place in a first-year writing course that is required of all students at the university. Anecdotally, most students are enrolled in these courses merely because they are required, so lack of volition (both VP and VC) is a significant issue. One problem is that many students do not perceive these courses as useful for their intended professions (Grau, 1999; Guberman, 2021). This problem is especially pervasive at open enrollment universities, like the one where the study took place. Internal survey data from the university of interest demonstrated that over 76% percent of students strongly agreed or somewhat agreed that they would not enroll in general education courses if they were not required. Only 30% of respondents strongly agreed or somewhat agreed that general education courses were relevant to their majors or intended professions. Many students dislike general education courses in general and liberal arts courses in particular. As both Guberman (2021) and Spidell and Thelin (2006) have explained, humanities courses are particularly disfavored by students. Accordingly, students enrolled in the RAW course may have had decreased volition from T<sub>1</sub> to T<sub>3</sub> (RQ2) for reasons that were unrelated to assessment practices.

It also is possible that ungrading has varying effects, depending on the characteristics of the student population. The study took place at a large, regional, open-enrollment, public university, whereas much of the research into ungrading strategies has occurred in graduate programs, highly selective universities, and high-performing students (e.g., Bloodgood et al., 2009; Meinking & Hall, 2022; Rohe et al., 2006; Seligman et al., 2021). Students at these universities are already highly motivated, as evidenced by the exemplary grades and tests scores required for admission to these universities. Perhaps the sample of participants surveyed for this study differed from students in other disciplines or at other universities in ways that decreased total volition.

**Research Question 3: How does student *perceived learning* (cognitive, affective, and psychomotor) change over time when ungrading is used in online, undergraduate RAW courses?**

Research question three examined changes in perceived learning over time when ungrading is used in undergraduate, online RAW courses. The results indicated a statistically significant decrease in perceived learning when ungrading was used. For the total CAP Scale, scores consistently decreased from T<sub>1</sub> to T<sub>3</sub> (RQ3). For the cognitive subscale (RQ3.1), scores at T<sub>3</sub> were higher than scores at T<sub>1</sub> (increased from T<sub>1</sub> to T<sub>2</sub> and decreased from T<sub>2</sub> to T<sub>3</sub>). For the affective subscale (RQ3.2), scores were lower at T<sub>3</sub> than at T<sub>1</sub> (decreased from T<sub>1</sub> to T<sub>2</sub> and increased from T<sub>2</sub> to T<sub>3</sub>). Finally, scores on the psychomotor subscale (RQ3.3) decreased steadily from T<sub>1</sub> to T<sub>3</sub>.

The findings are noteworthy, as prior research has indicated that ungrading increases actual and perceived learning (RQ3), while fostering a learning orientation rather than an achievement orientation (Hiller and Hietapelto, 2001; Inman & Powell,

2018; Lindemann & Harbke, 2011). Conversely, traditional grading often is associated with an achievement orientation, which largely results in surface learning (Ferguson, 2013; Tippin et al., 2012). For example, Pollio and Beck (2000) reported that their students were unable to adopt a learning orientation because they felt coerced by traditional grading practices that prioritize achievement over learning. Similarly, Chen and Zhang (2017) and Stuckey et al. (2020) found that students in ungraded courses reported increased levels of perceived learning. Despite these positive findings, others have argued that ungrading leads to complacency among both students and instructors. For example, McMorran and Ragupathi (2020) found that ungrading led to decreased performance and teaching to the weakest students rather than challenging students to excel. In the present study, it is possible that ungrading led to lack of academic rigor or complacency among both students and faculty, which might have reduced perceived learning (RQ3).

Lack of experience with ungrading also could have impacted perceived learning (RQ3). Some students prefer traditional grades because they are a known entity. Research supports this position. For example, research indicates that many students simply prefer grades because they are familiar (Brilleslyper et al., 2012; Cowan, 2020; Ferguson, 2016; Inman & Powell, 2018; Reardon and Guardado-Menjivar, 2020). Most students have participated in traditionally-graded courses throughout their educational experiences, and ungrading can be disorienting and uncomfortable for them, especially at the beginning of the semester before any assignments have been evaluated (Brilleslyper et al., 2012; Ferguson, 2016; Spidell & Thelin, 2006). In addition to students being unfamiliar with ungrading, both instructors were relatively new to ungrading, which may have minimized

the effectiveness of the intervention. Moreover, Keller and Suzuki (2004) have argued that clearly tying instruction to the learners' personal goals is critical. This factor is especially important in ungraded courses, as students must be motivated to participate in the absence of grades. A key struggle is to align all forms of assessment with learning goals and the learning process and to help students use assessments as opportunities for growth. It is possible that the instructors in this study failed to clearly connect the course material to students' personal learning goals, which reduced perceived learning.

### **Implications**

The present study was novel in that it sought to quantify the relationships among ungrading, motivation, volition, and perceived learning. More quantitative studies into ungrading are needed, as most of the existing studies are qualitative (e.g., Brilleslyper et al., 2012; Ferguson, 2013; Gomes et al., 2020; Hiller & Hietapelto, 2001; Inoue, 2020; Laflen & Sims, 2021; Mallette & Hawks, 2020; Reardon & Guardado-Menjivar, 2020; Reichert, 2003). This is especially true in liberal arts courses, such as RAW. Composition scholars have rarely investigated ungrading statistically, and much of the existing research is situated in the subjective experiences of individual instructors and their students (Cowan, 2020). Guberman (2021) also highlighted the need for more quantitative studies into ungrading, citing a need for evidence of its effectiveness. The present study has significant implications for educators, as it provides a more complete quantitative understanding of the effectiveness and limitations of ungrading practices.

The study also identified recommendations for practice. First, because the ungrading environment might be new to some students, instructors should provide explicit instruction regarding ungrading so that students will know what to expect and

will understand why ungrading is being used in the course (DeLuca et al., 2018; Horstmanshof & Brownie, 2013; Lam, 2018; Lee & Coniam, 2013). For example, students should be prepared for extensive formative feedback, as some students might perceive it to be overwhelming, which can reduce motivation, volition, and perceived learning. Students also should be coached in interpreting formative feedback so that they are better able to measure their learning progress. When students are new to ungrading, they can experience uncertainty and confusion regarding their performance, which can lead to decreased motivation, volition, and perceived learning. As Lam (2018) argues, students must “explicitly learn and then acquire the language of reflection and meta-discourse” (p. 30). This requires that the goals and success criteria must be clear to both students and instructors (Brookhart, 2001; DeLuca et al., 2018; Hawe & Parr, 2014; Lee & Coniam, 2013; Lipnevich et al., 2014; Regan, 2010).

In addition to coaching students about ungrading, instructors who adopt this assessment method should be trained in ungrading best practices (Lam, 2018; Lee & Coniam, 2013; Wiliam, 2006). Unless skillfully implemented, ungrading may be perceived as ambiguous or discouraging, which defeats its purpose. This concern is especially evident in online courses where there is an inherent physical barrier between the students and the instructor. Assessment can be challenging in all courses, but it is especially difficult in online courses. The present research suggests that ungrading might be more difficult to implement in the online environment. Therefore, instructors must ensure that ungrading will be effective for their demographic of students and in their academic subjects. Accordingly, instructors should be trained in ungrading best practices before incorporating these strategies in their courses.



Finally, although students received extensive formative feedback, assignments were graded using a binary grading scheme (i.e., *accept* or *reject*). Because instructors at the university of interest are required to assign grades, the range of ungrading strategies was limited. It is possible that a more nuanced ungrading strategy, such as negotiated grades or grading contracts, might be more effective at increasing motivation. It also might be helpful to add learning incentives. For example, Alt (2023) found that badges can be effective tools for some students when used in higher education, noting that badges alone are less effective than badges used in conjunction with other assessment methods (Alt 2023). It is possible that ungrading works best when accompanied by other assessment strategies.

### **Limitations and Recommendations**

This study had several limitations related to the sampling procedures. The study was conducted at one university using students from seven sections of RAW taught by two instructors, which represents a threat to external validity, as the results may not generalize to other settings. Additional research could explore ungrading with a more heterogeneous sample or in different academic disciplines. The sample size also was an additional potential limitation. The study included 57 participants, and the a priori power analysis indicated that a sample size of 61 was needed to test the hypotheses at 80% power with a medium effect size and a statistical significance criterion of  $\alpha = .05$ . Therefore, the study was slightly underpowered. A larger sample would be more representative of the population and would provide more valid and reliable results. The sampling method also represents a limitation. A convenience sample was used, and participants were divided by course section rather than using random sampling. It would

be helpful to replicate the study using a randomized sample rather than a convenience sample. Finally, because two instructors taught the RAW courses, differences in teaching styles, personalities, and grading practices may have impacted the study results.

Relatedly, the researcher taught three sections of the course, so there was the potential for response bias. Thus, future studies could use a random sample of participants taught by faculty members who are not primary researchers.

The surveying method also represents a potential limitation. Participants completed each instrument three times during the semester, which was a threat to validity, as familiarity with the instruments may have affected scores on the second and third administrations. Also, given the lower response rates at T<sub>2</sub> and especially at T<sub>3</sub>, it appears that some students might have experienced survey fatigue. Finally, because self-report instruments were used, the data could have been biased by false reporting or by student misjudgments regarding their levels of motivation (attention, relevance, confidence, and satisfaction), volition, and perceived learning. In this study, the surveys were administered in close succession. Future researchers could vary the timing of the surveys to mitigate these potential weaknesses.

The methodology of the study also might have been a limitation. This study used a repeated-measures design. Students completed the CIS, VFLS, and CAP Scale three times in an eight-week period. The results showed no statistically significant change in motivation when ungrading was used in undergraduate, online RAW courses.

Additionally, volition and perceived learning decreased from T<sub>1</sub> to T<sub>3</sub>. Research suggests that motivation may be particularly resistant to change to change (An and Li, 2021; Foong et al., 2022; Sulz et al., 2020). Therefore, it might be helpful to collect data over a

longer time period. Finally, the reliability of the cognitive subscale on the CAP Scale was quite low and  $T_2$  ( $\alpha = .363$ ) and  $T_3$  ( $\alpha = .305$ ). The subscale was retained for completeness of results. However, the cognitive subscale results may not be reliable.

Several areas for future research extend from the limitations and results of this study. First, because this study presents contradictory findings, it would be helpful to replicate the study. It is possible that this study is an anomaly and that future studies will achieve different results. Also, quantitative studies into the effectiveness of ungrading in first-language composition courses are limited. Additional studies situated in these courses would add to the literature. The literature review revealed no studies that have examined ungrading in the context of the ARCS model, so additional research that aligns these constructs could be valuable. Finally, the results of this study suggest that the effectiveness of ungrading could vary depending on the academic subject or delivery format. It is possible that ungrading is less effective in composition courses or in online courses, which has implications for educational practice. Studies incorporating ungrading in a variety of courses and academic disciplines would help to refine best practices, as instructors should choose high-impact interventions.

Future researchers also might compare the survey results with other measures of learning. For example, self-assessment helps students to develop the capacity for metacognition by encouraging them to analyze their progress towards learning objectives (Black et al., 2004; Reinholz, 2016). Peer-assessment can complement self-assessment because it provides practice in evaluating assignments and providing constructive feedback, skills that students can use to assess their own work (Brookhart, 2001; Harland et al., 2017; Hawe & Parr, 2014; Ibarra-Saiz et al. 2020; Lam, 2018; Regan, 2010). These

types of assessments could be added to ungrading, which might increase motivation, volition, and perceived learning. Future research could explore the viability of these types of alternative assessment.

The effectiveness of self-report measures of perceived learning also provides an area for future research, as students might have been unable to accurately self-assess their level of learning using the CAP Scale. Research suggests that students struggle to accurately quantify how much they have learned (Almohaimede, 2022; Teo et al., 2023; Yates et al, 2022). Because the CAP Scale is a self-report measure, it is possible that students were unable to accurately quantify their perceived learning. For example, some studies have found that students struggle to assess their own learning and that they often default to assessing their progress toward the instructor's goals (Kohn, 2011). It is, therefore, possible that students miscalculated their level of learning. Future researchers might explore this possibility by comparing students' perceived learning with tangible artifacts of their learning, such as essays and test scores.

### **Conclusion**

The purpose of this study was to examine changes in motivation, volition, and perceived learning when ungrading is used in online, undergraduate RAW courses. A quantitative, survey-based, repeated measures, correlational study was used to measure changes in these constructs over time. This study contradicts existing research into ungrading, an inconsistency that has practical implications. A critical review of the research literature suggests that students consistently demonstrate increased motivation, volition, and learning when ungrading practices are employed (Bloodgood et al., 2009; Brubaker, 2010; Bullock et al. 2022; Gorichanaz, 2022; Lindemann & Harbke, 2011;

Pulfrey et al., 2011; Rohe et al., 2006; Seligman et al. 2021). Prior studies suggest that students in ungraded classrooms are learning oriented rather than achievement oriented, which can increase perceived learning (Brubaker, 2010; Chamberlin et al., 2018; Ferguson, 2013; Gorichanaz, 2022; Grau, 1999; Guberman, 2021; McMorran & Ragupathi, 2020; Potts, 2010) They also have increased volition and are more willing to take academic risks (Cowan, 2020; Ferguson, 2013; Gorichanaz, 2022; Guberman, 2021; Mallette & Hawks, 2020; McMorran et al., 2017; McMorran & Ragupathi, 2020). Finally, ungrading practices are strongly correlated with improved learning outcomes (Chamberlin et al., 2018; Gorichanaz, 2022). However, the potential benefits of ungrading for increasing motivation, volition, and perceived learning were not realized in this study. The results suggest that ungrading might be less effective in certain types of courses or with certain types of students. It is possible, for example, that ungrading is more impactful in upper-division or graduate composition courses than in general education composition courses. It also might be true that ungrading is less effective in online courses. Therefore, educators who adopt ungrading should be mindful of its potential limitations.

## REFERENCES

- Adler-Kassner, L., & O'Neill, P. (2010). *Reframing writing assessment to improve teaching and learning*. Utah State University Press.
- Akbaş, O., Baturay, M., & Söker, Y. (2016). Interactive Whiteboard-Based ATM Use Training for Older Individuals. *International Online Journal of Educational Sciences*, 8(1), 87–97.  
[https://iojes.net/?mod=makale\\_ing\\_ozet&makale\\_id=40879](https://iojes.net/?mod=makale_ing_ozet&makale_id=40879)
- Alhassan, R. (2014). The effect of project-based learning and the ARCS motivational model on students' achievement and motivation to acquire database program skills. (2014). *Journal of Education and Practice*, 5(21), 158-164.  
<https://www.iiste.org/Journals/index.php/JEP/article/view/14528>
- Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *Clearing House*, 78(5), 218–223.  
<https://doi.org/10.3200/TCHS.78.5.218-223>
- Almohaimede, A. A. (2022). Comparison between students' self-evaluation and faculty members' evaluation in a clinical endodontic course at King Saud University. *European Journal of Dental Education*, 26(3), 569–576.  
<https://doi.org/10.1111/eje.12733>
- Alt, D. (2023). Who benefits from digital badges? Motivational precursors of digital badge usages in higher education. *Current Psychology*, 42(8), 6629-6640.  
<https://doi.org/10.1007/s12144-021-02002-0>
- An, M., & Li, L. (2021). The strength of motivation for medical school: A three-year longitudinal study. *Medical Teacher*, 43(9), 1079–1084.  
<https://doi.org/10.1080/0142159X.2021.1917765>
- Anderson, T., Rourke, L., Garrison, D.R., and Archer, W. (2001). Assessing social presence in asynchronous text-based computer conferencing. *Journal of Distance Education*, 14 (2), 1-17. <http://dx.doi.org/10.24059/olj.v5i2.1875>
- Angelo, T. (2017). Assessing motivation to improve learning: Practical applications of Keller's MVP model and ARCS-V design process. *New Directions for Teaching and Learning*, 152, 99-108. <http://dx.doi.org/10.1002/tl.20272>
- Annamalai, S. (2016). Implementing ARCS model to design a motivating multimedia e-book for polytechnic ESL classroom. *Journal of Telecommunication, Electronic and Computer Engineering*, 8(8), 57–60. <http://repo.uum.edu.my/id/eprint/20498/>
- Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Press.

- Barber, M. (2021). Making the grade: Canadian students' attitudes regarding assessment. *The International Journal of Learning in Higher Education*, 28(1), 201-206. <https://doi.org/10.18848/2327-7955/CGP/v28i01/201-206>
- Bawa, P. (2016). Retention in online courses. *SAGE Open*, 6(1), 1-11. <https://doi.org/10.1177/2158244015621777>
- Berge, Z.L. and Collins, M. (1995). (Eds.) *Computer-mediated communication and the Online classroom*. Hampton Press.
- Black, P., Harrison, C., Lee, C., Marshall, B., Wiliam, D. (2003). *Assessment for Learning: Putting It into Practice*. Open University Press.
- Black, P., Harrison, C., Lee, C., Marshall, B., Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *The Phi Delta Kappan*, 86(1), 8-21. <https://doi.org/10.1177/003172170408600105>
- Black, P. & Wiliam, D. (1998) Assessment and classroom learning: Principles, policy, and practice. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability*, 21, 5-31. <https://doi.org/10.1007/s11092008-9068-5>
- Bloodgood, R. A., Short, J. G., Jackson, J. M., & Martindale, J. R. (2009). A change to pass/fail grading in the first two years at one medical school results in improved psychological well-being. *Academic Medicine*, 84(5), 655–662. <https://doi.org/10.1097/ACM.0b013e31819f6d78>
- Bloom, B. S. (1971). *Mastery learning*. Holt, Rineheart, & Winston.
- Blum, S. (2020). *Ungrading: Why Rating Students Undermines Learning (and What to Do Instead)*. West Virginia University Press.
- Borg, E. (2003). Discourse Community. *ELT Journal*, 57(4), 398–400. <https://doi.org/10.1093/elt/57.4.398>
- Brilleslyper, M., Ghrist, M., Holcomb, T., Schaubroeck, B., Warner, B., & Williams, S. (2012). What's the point? The benefits of grading without points. *PRIMUS*, 22(5), 411–427. <https://doi.org/10.1080/10511970.2011.571346>
- Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice*, 8(2), 153–169. <https://doi.org/10.1080/09695940123775>

- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research*, 86(4), 803-848. <https://doi.org/10.7916/D8NV9JQ0>
- Brubaker, N. D. (2010). Negotiating authority by designing individualized grading contracts. *Studying Teacher Education*, 6(3), 257-267. <https://doi.org/10.1080/17425964.2010.518667>
- Bullock, J. L., Seligman, L., Lai, C. J., O'Sullivan, P. S., & Hauer, K. E. (2022). Moving toward mastery: Changes in student perceptions of clerkship assessment with pass/fail grading and enhanced feedback. *Teaching and Learning in Medicine*, 34(2), 198–208. <https://doi.org/10.1080/10401334.2021.1922285>
- Burger, R. (2017). Student perceptions of the fairness of grading procedures: A multilevel investigation of the role of the academic environment. *Higher Education*, 72(2), 301-320. <https://doi.org/10.1007/s10734-016-0049-1>
- Burton, G. E., Pathak, D. S., & Burton, D. B. (1977). Effects of contract grading on student evaluations of teachers. *Educational Horizons* 55(3), 128–133. <https://www.jstor.org/stable/42924231>
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, 79(4), 474-482. <https://doi.org/10.1037/0022-0663.79.4.474>
- Canfield, M. L., Kivisalu, T. M., van der Karr, C., King, C., & Phillips, C. E. (2015). The use of course grades in the assessment of student learning outcomes for general education. *SAGE Open*, 5(4), 1-13. <https://doi.org/10.1177/2158244015615921>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <http://10.0.4.56/02602938.2018.1463354>
- Chamberlin, K., Yasue, M., & I-Chant, A. C. (2018). The impact of grades on student motivation. *Active Learning in Higher Education*, 1-16. <https://doi.org/10.1177/1469787418819728>
- Chang, C., Chang, C. K., & Shih, J. L. (2016). Motivational strategies in a mobile inquiry-based language learning setting. *System*, 59, 100–115. <https://doi.org/10.1016/j.system.2016.04.013>
- Chang, M. M., & Lehman, J. D. (2002). Learning foreign language through an interactive multimedia program: An experimental study on the effects of the relevance



- component of the ARCS model. *Calico Journal*, 20(1), 81–98.  
<https://www.jstor.org/stable/24149610>
- ChanMim, K., Park, S. W., Huynh, N., & Schuermann, R. T. (2017). University students' motivation, engagement, and performance in a large lecture-format general education course. *Journal of Further and Higher Education*, 41(2), 201-214.  
<http://dx.doi.org/10.1080/0309877X.2015.1070401>.
- Chao, J., Jiang, T. W., Yeh, Y. H., Liu, C. H., & Lin, C. M. (2019). Integration of ARCS motivational model and IT to enhance students learning in the context of Atayal culture. *EURASIA Journal of Mathematics, Science and Technology Education*, 15(11), Article em1771. <https://doi.org/10.29333/ejmste/109608>
- Chen, D., & Zhang, L. (2017). Formative assessment of academic English writing for Chinese EFL learners. *TESOL International Journal*, 12(2), 47-64.  
<https://eric.ed.gov/?id=EJ1247811>
- Chowdhury, F. (2018). Grade inflation: Causes, consequences, and cure. *Journal of Education and Learning*, 7(6), 86–92. <https://doi.org/10.5539/jel.v7n6p86>
- Cisek, G. J. (2000). Pockets of resistance in the assessment revolution. *Educational Measurement: Issues and Practice*, 19(2), 16-23. <https://doi.org/10.1111/j.1745-3992.2000.tb00026.x>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.)  
Routledge.
- Cowan, M. (2020). A legacy of grading contracts for composition. *Journal of Writing Assessment* 13(2), Article 150.  
<http://journalofwritingassessment.org/article.php?article=150>
- Çolakoğlu, Ö., & Omur, A. (2010). Motivational measure of the instruction compared: Instruction based on the ARCS motivation theory vs. traditional instruction in blended courses. *The Turkish Online Journal of Distance Education*, 11(2), Article 3. <https://eric.ed.gov/?id=EJ1042575>
- Connors, R. J., & Lunsford, A. A. (1993). Teachers' rhetorical comments on student papers. *College Composition and Communication*, 44(2), 200-223.  
<https://www.jstor.org/stable/358839>
- Cowan, M. (2020). A legacy of grading contracts for composition. *Journal of Writing Assessment*, 13(2), Article 150.  
<http://journalofwritingassessment.org/article.php?article=150>
- Danielewicz, J., & Elbow, P. (2009). A unilateral grading contract to improve learning

- and teaching. *College Composition & Communication*, 61(2), 244–268.  
<https://www.jstor.org/stable/40593442>
- Deci, E., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum.
- DeLuca, C., Chapman-Chin, A. E. A., LaPointe-McEwan, D., & Klinger, D. A. (2018). Student perspectives on assessment for learning. *Curriculum Journal*, 29(1), 77–94. <https://doi.org/10.1080/09585176.2017.1401550>
- Denton, P., Madden, J., Roberts, M., & Rowe, P. (2008). Students' response to traditional and computer-assisted formative feedback: A comparative case study. *British Journal of Educational Technology* 39(3), 486-500.  
<https://doi.org/10.1111/j.1467-8535.2007.00745.x>
- Denton, P., & Rowe, P. (2015). Using statement banks to return online feedback: Limitations of the transmission approach in a credit-bearing assessment. *Assessment and Evaluation in Higher Education*, 40(8), 1095–1103.  
<https://doi.org/10.1080/02602938.2014.970124>
- Douglas, K. A., Merzdorf, H. E., Hicks, N. M., Sarfraz, M. I., & Bermel, P. (2020). Challenges to assessing motivation in MOOC learners: An application of an argument-based approach. *Computers & Education*, 150, Article 103829.  
<https://doi.org/10.1016/j.compedu.2020.103829>
- Dweck, C. (2016). *Mindset: The new psychology of success* (updated edition). Ballantine.
- Elbow, P. (1968). A method for teaching writing. *College English*, 30(2), 115-125.  
<https://doi.org/10.2307/374447>
- Elbow, P. (1983). Embracing contraries in the teaching process. *College English*, 45(4), 327-339. <https://doi.org/10.2307/376541>
- Elbow, P. (1993). Ranking, evaluating, and liking: Sorting out three forms of judgment. *College English*, 55(2), 187-206. <https://doi.org/10.2307/378503>
- Elbow, P. (1997a). Grading student writing: Making it simpler, fairer, clearer. *New Directions for Teaching and Learning*, 69, 127–140.  
<https://doi.org/10.1002/tl.6911>
- Elbow, P. (1997b). Taking time out from grading and evaluating while working in a conventional system. *Assessing Writing*, 4(1), 5-27.  
[https://doi.org/10.1016/S1075-2935\(97\)80003-7](https://doi.org/10.1016/S1075-2935(97)80003-7)
- Elkins, D. M. (2016). Grading to learn: An analysis of the importance and application of

specifications grading in a communication course. *Kentucky Journal of Communication*, 35(2), 26–48.

- Ferguson, H. J. (2013). Journey into ungrading. *Counterpoints*, 451, 194-209.  
<http://www.jstor.org/stable/42982092>
- Firat, M., Kilinc, H., & Yuzer, V. (2018). Level of intrinsic motivation of distance education students in e-learning environments. *Journal of Computer Assisted Learning*, 34, 63-70. <https://doi.org/10.1111/jcal.12214>
- Foong, C. C., Liew, P. Y., & Lye, A. J. (2022). Changes in motivation and its relationship with academic performance among first-year chemical engineering students. *Education for Chemical Engineers*, 38, 70–77.  
<https://doi.org/10.1016/j.ece.2021.11.002>
- Franklin, J. (2017). MVP and instructional systems design in online courses. *New Directions for Teaching and Learning*, 152, 39-52.  
<https://doi.org/10.1002/tl.20267>
- Gabrielle, D. (2003). The effects of technology-mediated instructional strategies on motivation, performance, and self-directed learning. [Doctoral dissertation, Florida State University]. ProQuest Dissertations and Theses Global.
- Garrison, D. R., & Anderson, T. (2003). *E-learning in the 21st century*. Routledge.
- Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education*, 2(2-3), 87-105.  
[https://doi.org/10.1016/S1096-7516\(00\)00016-6](https://doi.org/10.1016/S1096-7516(00)00016-6)
- Goksu, I., & Bolat, Y. I. (2021). Does the ARCS motivational model affect students' achievement and motivation? A meta-analysis. *Review of Education*, 9(1), 27–52.  
<https://doi.org/10.1002/rev3.3231>
- Gomes, M. (2020). Enabling meaningful labor: Narratives of participation in a grading contract. *Journal of Writing Assessment*, 13(2), Article 152.  
<http://journalofwritingassessment.org/article.php?article=152>
- Gorichanaz, T. (2022). “It made me feel like it was okay to be wrong”: Student experiences with ungrading. *Active Learning in Higher Education*, 1-13.  
<https://doi.org/10.1177/14697874221093640>
- Grau, H. J. (1999). “Streamlined” contract grading – A performance-measuring

- alternative to traditional evaluation methods: Raising the importance of concept mastery by reducing the emphasis on grades. *Journal of College Science Teaching*, 28(4), 254–258. <https://www.jstor.org/stable/42990676>
- Gruhlke, H. (2018). The ethics of grade inflation: Investigating the ethical implications of grade inflation in U.S. colleges. *Global Education Journal*, 2018(2), 1-17.
- Guberman, D. (2021). Student perceptions of an online ungraded class. *Teaching and Learning Inquiry*, 9(1), 86-98. <https://doi.org/10.20343/teachlearninqu.9.1.8>
- Hao, K. C. & Lee, L. C. (2019) The development and evaluation of an educational game integrating augmented reality, ARCS model, and types of games for English experiment learning: an analysis of learning, *Interactive Learning Environments*, 29(7), 1101–1114, <https://doi.org/10.1080/10494820.2019.1619590>
- Harland, T., Wald, N., & Randhawa, H. (2017). Student peer review: Enhancing formative feedback with a rebuttal. *Assessment & Evaluation in Higher Education*, 42(5), 801–811. <https://doi.org/10.1080/02602938.2016.1194368>
- Harrison, C. J., Konings, K. D., Schuwirth, L., Wass, V., & Vleuten, C. (2015). Barriers to the uptake and use of feedback in the context of summative assessment. *Advances in Health Sciences Education*, 20, 229-245. <https://doi.org/10.1007/s10459-014-9524-6>
- Hattie, J., & Timperley, F. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Hawe, E., & Parr, J. (2014). Assessment for learning in the writing classroom: An incomplete realisation. *The Curriculum Journal*, 25(2), 210-237. <https://doi.org/10.1080/09585176.2013.862172>
- Hiller, T. B., & Hietapelto, A. B. (2001). Contract grading: Encouraging commitment to the learning process through voice in the evaluation process. *Journal of Management Education*, 25(6), 660–684. <https://doi.org/10.1177/105256290102500605>
- Hobson, T. D., & Puruhito, K. K. (2018). Going the distance: Online course performance and motivation of distance-learning students. *Online Learning*, 22(4), 129-140. <http://dx.doi.org/10.24059/olj.v22i4.1516>
- Horstmanshof, L., & Brownie, S. (2013). A scaffolded approach to discussion board use for formative assessment of academic writing skills. *Assessment & Evaluation in Higher Education*, 38(1), 61-73. <https://doi.org/10.1080/02602938.2011.604121>
- Huett, J. B. (2006). The effects of ARCS-based confidence strategies on learner

confidence and performance in distance education. [Unpublished doctoral dissertation]. University of North Texas.

- Huett, J. B., Moller, L., Young, J., Bray, M., & Huett, K. C. (2008). Supporting the distant student: the effect of ARCS-based strategies on confidence and performance. *Quarterly Review of Distance Education*, 9(2), 113-126. <https://eric.ed.gov/?id=EJ875091>
- Hung, I. C., Chao, K. J., Lee, L., & Chen, N. S. (2013). Designing a robot teaching assistant for enhancing and sustaining learning motivation. *Interactive Learning Environments*, 21(2), 156–171. <https://doi.org/10.1080/10494820.2012.705855>
- Huot, B. A. (2002). *(Re)articulating writing assessment for teaching and learning*. Utah State University Press.
- Iamarino, D. L. (2014). The benefits of standards-based grading: A critical evaluation of modern grading practices. *Current Issues in Education*, 17(2), 1-12. <https://cie.asu.edu/ojs/index.php/cieatasu/article/download/1234/568/>
- Ibarra-Sáiz, M.S., Rodríguez-Gómez, G. & Boud, D. (2020). Developing student competence through peer assessment: The role of feedback, self-regulation, and evaluative judgement. *Higher Education*, 80, 137–156. <https://doi.org/10.1007/s10734-019-00469-2>
- Inman, J. O., & Powell, R. A. (2018). In the absence of grades: Dissonance and desire in course-contract classrooms. *College Composition and Communication*, 70(1), 30–56. <https://www.jstor.org/stable/26772544>
- Inoue, A. B. (2004). Community-based assessment pedagogy. *Assessing Writing*, 9(3), 208–238. <https://doi.org/10.1016/j.asw.2004.12.001>
- Inoue, A. B. (2012a). Grading contracts: Assessing their effectiveness on different racial formations. In A. B. Inoue & M. Poe (Eds.), *Race and writing assessment* (pp. 78–93). Peter Lang.
- Inoue, A. B. (2012b). Racial methodologies for composition studies: Reflecting on theories of race in writing assessment research. In L. Nickoson & M. P. Sheridan (Eds.), *Writing studies research in practice: Methods and methodologies* (pp. 125–139). Southern Illinois University Press.
- Inoue, A. B. (2014). A grade-less writing course that focuses on labor and assessing. In D. Coxwell-Teague & R. Lunsford (Eds.), *First-year composition: From theory to practice* (pp. 71–110). Parlor Press.
- Inoue, A. B. (2019). Labor-based grading contracts: Building equity and inclusion in the compassionate writing classroom. University Press of Colorado.

- Inoue, A. B. (2020). Stories about grading contracts, or how do I like through the violence I've done? *Journal of Writing Assessment*, 13(2), Article 154. <http://journalofwritingassessment.org/article.php?article=154>
- Jeon, E. Y. (2021). What makes them the best English teachers? An analysis of the motivational strategy use based on ARCS model. *Educational Research for Policy and Practice*, 20(3), 263-278. <https://doi.org/10.1007/s10671-020-09279-1>
- Jeon, K., Jarrett, O. S., & Ghim, H. D. (2014). Project-based learning in engineering education: Is it motivational? *International Journal of Engineering Education*, 30(2), 438–448. <https://www.researchgate.net/publication/285579124>
- Jokelova, A. (2013). ARCS motivational model: Theoretical concepts and its use in online courses. *2013 IEEE 11th International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 189–194. <https://doi.org/10.1109/ICETA.2013.6674427>
- Karakış, H., Karamete, A., & Okçu, A. (2016). The effects of a computer-assisted teaching material, designed according to the ASSURE instructional design and the ARCS model of motivation, on students' achievement levels in a mathematics lesson and their resulting attitudes. *European Journal of Contemporary Education*, 15(1), 105–113. <https://doi.org/10.13187/ejced.2016.15.105>
- Kearsley, G. (2002). Is online learning for everybody? *Educational Technology*, 42(1), 41-44. <https://www.jstor.org/stable/44428721>
- Keller, J. M. (1979). Motivation and instructional design: A theoretical perspective. *Journal of Instructional Development* 2(4), 26-24. <https://www.jstor.org/stable/30220576>
- Keller, J. M. (1983). Motivational design of instruction. In C. M. Reigeluth (Ed.), *Instructional design theories and models: A overview of their current status* (pp. 383-433). Lawrence Erlbaum Associates.
- Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development*, 10(3), 2-10. <https://www.jstor.org/stable/30221294>
- Keller, J. M. (2008a). An integrative theory of motivation, volition, and performance. *Technology, Instruction, Cognition, and Learning* 6(2), 79-104. <http://terrikrause.com/Content/documents/Keller2008IntegrativeTheory.pdf>
- Keller, J. M. (2008b). First principles of motivation to learn and e-learning. *Distance Education*, 29(2), 175-185. <https://doi.org/10.1080/01587910802154970>
- Keller, J. M. (2010). *Motivational design for learning and performance: the ARCS model*

*approach*. Springer.

- Keller, J. M. (2016). Motivation, learning, and technology: Applying the ARCS motivation model. *Participatory Educational Research*, 3(2), 1-13. <http://dx.doi.org/10.17275/per.16.06.3.2>
- Keller, J. M. (2017). The MVP model: Overview and application. *New Directions for Teaching and Learning*, 152, 13-26. <https://doi.org/10.1002/tl.20265>
- Keller, J. M., & Suzuki, K. (1988). Use of the ARCS motivation model in courseware design. In D. H. Jonassen (Ed.), *Instructional designs for micro-computer courseware* (pp. 401-434). Erlbaum.
- Keller, J. M., & Suzuki, K. (2004). Learner motivation and e-learning design: A multinationally validated process. *Journal of Educational Media*, 29(3), 229-239. <https://doi.org/10.1080/1358165042000283084>
- Keller, J. M., Ucar, H., & Kumtepe, A. T. (2020). Development and validation of a scale to measure volition for learning. *Open Praxis*, 12(2), 161–174. <https://doi.org/10.5944/openpraxis.12.2.1082>
- Koenka, A. C. (2020). Grade expectations: The motivational consequences of performance feedback on a summative assessment. *The Journal of Experimental Education*, 90(1), 88-111. <https://doi.org/10.1080/00220973.2020.1777069>
- Koenka, A. C., Linnenbrink-Garcia, L. Moshontz, H., Atkinson, K. M., Sanchez, C. E., & Cooper, H. (2019). A meta-analysis on the impact of grades and comments on academic motivation and achievement: A case for written feedback. *Educational Psychology*, 41(7), 922-947. <https://doi.org/10.1080/01443410.2019.1659939>
- Kohn, A. (1994). Grading: The issue is not how but why. *Educational Leadership*, 52(2), 38-41.
- Kohn, A. (1999). D-grading is degrading. *Education Digest*, 65(1), 59.
- Kohn, A. (2002). The dangerous myth of grade inflation. *Chronicle of Higher Education*, 49(11), 1-10. <https://www.chronicle.com/article/the-dangerous-myth-of-grade-inflation/>
- Kohn, A. (2011). The case against grades. *Educational Leadership*, 69(3), 28–33. <https://www.jstor.org/stable/42982088>
- Kurt, P. Y., & Keçik, İ. (2017). The effects of ARCS motivational model on student motivation to learn English. *European Journal of Foreign Language Teaching*, 2(1), 22-44. <https://oapub.org/edu/index.php/ejfl/article/view/478>

- Laflen, A., & Sims, M. (2021). Designing a more equitable scorecard: Grading contracts and online writing instruction. In J. Borgman & C. McArdle (Eds.), *PARS in Practice: More Resources and Strategies for Online Writing Instructors* (pp. 119–139). WAC Clearinghouse.
- Lam, R. (2016). Assessment as learning: examining a cycle of teaching, learning, and assessment of writing in the portfolio-based classroom. *Studies in Higher Education, 41*(11), 1900–1917. <https://doi.org/10.1080/03075079.2014.999317>
- Lam, R. (2018). Understanding assessment as learning in writing classrooms: The case of portfolio assessment. *Iranian Journal of Language Teaching Research, 6*(3), 19–36. <https://doi.org/10.1080/03075079.2014.999317>
- Landrum, R. E., & Dietz, K. H. (2006). Grading without points: Does it hurt student performance? *College Teaching, 54*(4), 298–301. <https://www.jstor.org/stable/27559291>
- Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing, 22*, 34–50. <https://doi.org/10.1016/j.jslw.2012.11.003>
- Li, K., & Keller, J. M. (2018). Use of the ARCS model in education: A literature review. *Computers and Education, 122*, 54–62. <https://doi.org/10.1016/j.compedu.2018.03.019>
- Lindemann, D. F., & Harbke, C. R. (2011). Use of contract grading to improve grades among college freshmen in introductory psychology. *SAGE Open, 1*(3), 1–7. <https://doi.org/10.1177/2158244011434103>
- Lipnevich, A. A., McCallen, L. N., Miles, K. P., & Smith, J. K. (2014). Mind the gap! Students' use of exemplars and detailed rubrics as formative assessment. *Instructional Science, 42*(4), 539–559. <https://doi.org/10.1007/s11251-013-9299-9>
- Litterio, L. M. (2016). Contract grading in a technical writing classroom: A case study. *Journal of Writing Assessment, 9*(2), 1–8. <http://journalofwritingassessment.org/article.php?article=101>
- Litterio, L. M. (2018). Contract grading in the technical writing classroom: Blending community-based assessment and self-assessment. *Assessing Writing, 38*, 1–9. <https://doi.org/10.1016/j.asw.2018.06.002>
- Lynch, R., & Hennessy, J. (2017). Learning to earn? The role of performance grades in higher education. *Studies in Higher Education, 42*(9), 1750–1763. <https://doi.org/10.1080/03075079.2015.1124850>
- Lynn, P. (2004). *Coming to terms: A theory of writing assessment*. Utah State University



Press. [https://digitalcommons.usu.edu/usupress\\_pubs/149](https://digitalcommons.usu.edu/usupress_pubs/149)

- Mallette, J. C., & Hawks, A. (2020). Building student agency through contract grading in technical communication. *Journal of Writing Assessment*, 13(2), Article 158. <http://journalofwritingassessment.org/article.php?article=158>
- Marshall, D. B., Beck, S. L., Coghlan, L. K., & Kimbro, E. (2005). Understanding student learning outcomes through narrative transcript analysis: Assessing general education in an institution without grades or required courses. *Assessment Update*, 17(5), 4-11. <https://eric.ed.gov/?id=EJ790787>
- Martin, J. H., & Harrison, C. H. (1972). *Free to learn: Unlocking and ungrading American education*. Prentice-Hall.
- McCroskey, J. C. (1994). Assessment of affect toward communication and affect toward instruction in communication. In S. Morreale & M. Brooks (Eds.), *1994 SCA Summer Conference Proceedings and Prepared Remarks: Assessing College Student Competence in Speech Communication*. Annandale, VA: Speech Communication Association.
- McMorran, C., Luo, S., & Ragupathi, K. (2017). Assessment and learning without grades? Motivations and concerns with implementing gradeless learning in higher education. *Assessment and Evaluation in Higher Education*, 42(3), 361–377. <https://doi.org/10.1080/02602938.2015.1114584>
- McMorran, C., & Ragupathi, K. (2020). The promise and pitfalls of gradeless learning: Responses to an alternative approach to grading. *Journal of Further and Higher Education*, 44(7), 925–938. <https://doi.org/10.1080/0309877X.2019.1619073>
- Means, T. B., Jonassen, D. H., & Dwyer, F. M. (1997). Enhancing relevance: Embedded ARCS strategies vs. purpose. *Educational Technology Research & Development*, 45(1), 5–17. <https://doi.org/10.2307/30220166>
- Meinking, K., & Hall, E. E. (2022). Letting go of grades: Creating an environment of autonomy and a focus on learning for high achieving students. *Teaching & Learning Inquiry*, 10. <https://doi.org/10.20343/teachlearninqu.10.21>
- Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research & Development*, 50(3), 43-59. <https://doi.org/10.1007/BF02505024>
- Merva, M. (2003). Grades as incentives: A quantitative assessment with implications for study abroad programs. *Journal of Studies in International Education*, 7(2), 149-156. <https://doi.org/10.1177/1028315303007002003>
- Michaelides, M., & Kirshner, B. (2005). Graduate student attitudes toward grading systems. *College Quarterly*, 8(4). <https://eric.ed.gov/?id=EJ848740>

[http://collegequarterly.ca/2005-vol08-num04-fall/michaelides\\_kirshner.html](http://collegequarterly.ca/2005-vol08-num04-fall/michaelides_kirshner.html)

- Milman, N. B., & Wessmiller, J. (2020). Motivating the online learner using Keller's ARCS model. *Distance Learning*, 17(4), 33-37.
- Mirzaei, A., Rad, H. S., & Rahimi, E. (2022). Integrating ARCS motivational model and flipped teaching in L2 classrooms: A case of EFL expository writing. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2022.2068614>
- Morgan, G. A., Barrett, K. C., Leech, N. L., & Gloeckner, G. W. (2019). *IBM SPSS for introductory statistics: Use and interpretation* (5th ed.). Taylor and Francis.
- Moura, A., & Carvalho, A. A. (2012). The ARCS model to motivate language learning through SMS and podcasts. *Journal of Communications Research*, 4(4), 385–406.
- Murray, D. M. (2003). Teach writing as a process not product. In V. Villanueva (Ed.) *Cross-Talk in Comp Theory: A Reader*. (2nd ed., pp. 3-6). National Council of Teachers of English.  
<https://eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED474209>
- Naveed, T., & Bhatti, N. M. (2017). Validation of course interest survey tool in MBBS students of a private medical college of Pakistan. *Journal of Islamic International Medical College*, 12(4), 201-207. <https://www.bibliomed.org/?jtt=1815-4018&la=0&yr=2017&vol=12&issue=4.000>
- Neal, M. R. (2011). *Writing assessment and the revolution in digital texts and technologies*. Teachers College Press.
- Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <http://10.0.4.56/03075070600572090>
- Ninomiya, S. (2016). The possibilities and limitations of assessment for learning: Exploring the theory of formative assessment and the notion of “closing the learning gap.” *Educational Studies in Japan: International Yearbook*, 10, 79–91. <https://doi.org/10.7571/esjkyoiku.10.79>
- Novak, E. (2014). Toward a mathematical model of motivation, volition, and performance. *Computers and Education*, 74, 73-80.  
<http://dx.doi.org/10.1016/j.compedu.2014.01.009>
- Nystrom, K. (2018). When students are allowed to choose: Grading scale choices for degree projects. *Studies in Higher Education*, 43(11), 1993-2002.  
<https://doi.org/10.1080/03075079.2017.1290062>
- Pan, X., & Shao, H. (2020). Teacher online feedback and learning motivation: Learning

- engagement as a mediator. *Social Behavior & Personality: An International Journal*, 48(6), 1–10. <https://doi.org/10.2224/sbp.9118>
- Park, S., & Yun, H. (2017). Relationships between motivational strategies and cognitive learning in distance education courses. *Distance Education*, 38(3), 302–320. <https://doi.org/10.1080/01587919.2017.1369007>
- Penn, P., & Wells, I. (2017). Enhancing feedback and feed-forward via integrated virtual learning environment based evaluation and support. *Psychology Teaching Review*, 23(2), 60–65. <https://eric.ed.gov/?id=EJ1160271>
- Polczynski, J. J., & Shirland, I. E. (1977). Expectancy theory and contract grading combined as an effective motivational force for college students. *Journal of Educational Research*, 70(5), 238–241. <http://10.0.4.56/00220671.1977.10884996>
- Pollio, H. R., & Beck, H. P. (2000). When the tail wags the dog: Perceptions of learning and grade orientation in, and by, contemporary college students and faculty. *The Journal of Higher Education*, 71(1), 84–102. <https://doi.org/10.2307/2649283>
- Poorman, S. G., & Mastorovich, M. L. (2019). The meaning of grades: Stories of undergraduate, master’s, and doctoral nursing students. *Nurse Educator*, 44(6), 321–325. <https://doi.org/10.1097/NNE.0000000000000627>
- Potts, G. (2010). A Simple Alternative to Grading. *Inquiry*, 15(1), 29–42. <https://commons.vccs.edu/inquiry/vol15/iss1/4/>
- Proske, A., Roscoe, R., & McNamara, D. (2014). Game-based practice versus traditional practice in computer-based writing strategy training: effects on motivation and achievement. *Educational Technology Research & Development*, 62(5), 481–505. <https://doi.org/10.1007/s11423-014-9349-2>
- Pulfrey, C., Butera, F., & Buchs, C. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology*, 103(3), 683–700. <https://doi.org/10.1037/a0023911>
- Pulfrey, C., Darnon, C., & Butera, F. (2013). Autonomy and task performance: Explaining the impact of grades on intrinsic motivation. *Journal of Educational Psychology*, 105(1), 39–57. <https://doi.org/10.1037/a0029376>
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4–13. <https://doi.org/10.1002/bs.3830280103>
- Reddan, G. (2013). To grade or not to grade: Student perceptions of the effects of grading a course in work-integrated learning. *Asia-Pacific Journal of Cooperative Education*, 14(4), 223–232. <https://research->

[repository.griffith.edu.au/bitstream/handle/10072/49657/84670\\_1.pdf?sequence=1](http://repository.griffith.edu.au/bitstream/handle/10072/49657/84670_1.pdf?sequence=1)

- Reardon, K., & Guardado-Menjivar, V. (2020). Perceptions of fairness in summer bridge classrooms with contract grades. *Journal of Writing Assessment*, 13(2), Article 153. <http://journalofwritingassessment.org/article.php?article=153>
- Reinholz, D. (2018). Three approaches to focusing peer feedback. *International Journal for the Scholarship of Teaching and Learning*, 12(2), Article 10. <https://doi.org/10.20429/ijstl.2018.120210>
- Regan, P. (2010). Read between the lines: The emancipatory nature of formative annotative feedback on draft assignments. *Systemic Practice & Action Research*, 23(6), 453–466. <https://doi.org/10.1007/s11213-010-9168-2>
- Reichert, N. (2003). Practice makes perfect: Contracting quantity and quality. *Teaching English in the Two Year College*, 31(1), 60–69.
- Richmond, V. P., Gorham, J. S., & McCroskey, J. C. (1987). The relationship between selected immediacy behaviors and cognitive learning. In M. McLaughlin (Ed.), *Communication Yearbook 10* (pp. 188–198). Sage.
- Riddell, J. (2015). Performance, feedback, and revision: Metacognitive approaches to undergraduate essay writing. *Collected Essays on Learning & Teaching*, 8, 79–95. <http://10.0.87.57/celt.v8i0.4256>
- Rifkin, B. (2005). A ceiling effect in traditional classroom foreign language instruction: Data from Russian. *The Modern Language Journal*, 89(1), 3–18. <https://doi.org/10.1111/j.0026-7902.2005.00262.x>
- Roberts, R. M., & Dorstyn, D. S. (2017). Does changing grade-based marking to pass-fail marking change student achievement in a postgraduate psychology course. *Training and Education in Professional Psychology*, 11(1), 57–60. <https://doi.org/10.1037/tep0000129>
- Rockinson-Szapkiw, A. J., Pritchard, T., McComb-Beverage, S., & Schellenberg, R. (2013). Promoting professional identity: A within group comparison of wiki-based and traditional assignments on school counselling students' learning, sense of community and computer anxiety. *British Journal of Guidance and Counselling*, 41(5), 559–572. <http://dx.doi.org/10.1080/03069885.2013.773958>
- Rohe, D. E., Barrier, P. A., Clark, M. M., Cook, D. A., Vickers, K. S., & Decker, P. A. (2006). The benefits of pass-fail grading on stress, mood, and group cohesion in medical students. *Mayo Clinic Proceedings*, 81(11), 1443–1448.
- Rovai, A. P., Wighting, M. J., Baker, J. D., Grooms, L. D. (2009). Development of an

- instrument to measure perceived cognitive, affective, and psychomotor learning in traditional and virtual classroom higher education settings. *Internet and Higher Education*, 12, 7-13. <https://doi.org/10.1016/j.iheduc.2008.10.002>
- Sadler, D. R. (2009). Grade integrity and the representation of academic achievement. *Studies in Higher Education*, 34(7), 807–826. <https://doi.org/10.1080/03075070802706553>
- Sapp, D. A., & Simon, J. (2005). Comparing grades in online and face-to-face writing courses: Interpersonal accountability and institutional commitment. *Computers and Composition: An International Journal for Teachers of Writing*, 22(4), 471–489. <https://doi.org/10.1016/j.compcom.2005.08.005>
- Schinske, J., & Tanner, K. (2014). Teaching more by grading less (or differently). *CBE Life Sciences Education*, 13(2), 159–166. <https://doi.org/10.1187/cbe.CBE-14-03-0054>
- Schneider, J., & Hutt, E. (2014). Making the grade: a history of the A-F marking scheme. *Journal of Curriculum Studies*, 46(2), 201–224. <https://doi.org/10.1080/00220272.2013.790480>
- Sdravopoulou, K., Muñoz González, J. M., & Hidalgo-Ariza, M. D. (2021). Assessment of a location-based mobile augmented-reality game by adult users with the ARCS model. *Applied Sciences (2076-3417)*, 11(14), 6448. <https://doi.org/10.3390/app11146448>
- Seligman, L., Abdullahi, A., Teherani, A., & Hauer, K. E. (2021). From grading to assessment for learning: A qualitative study of student perceptions surrounding elimination of core clerkship grades and enhanced formative feedback. *Teaching & Learning in Medicine*, 33(3), 314–325. <https://doi.org/10.1080/10401334.2020.1847654>
- Shor, I. (1996). *When students have power: Negotiating authority in a critical pedagogy*. University of Chicago Press.
- Shor, I. (2009). Critical pedagogy is too big to fail. *Journal of Basic Writing*, 28(2), 6–27. <https://www.jstor.org/stable/43443879>
- Smith, J. A. (1999). Contracting English composition: It only sounds like an illness. *Teaching English in the Two Year College*, 26(4), 427-430.
- Soles, D. (2001). Grading as a teaching strategy. *Teaching English in the Two Year College*, 29(2), 122-134.
- Spidell, C., & Thelin, W. H. (2006). Not ready to let go: A study of resistance to grading

- contracts. *Composition Studies*, 34(1), 35-68.  
<https://www.jstor.org/stable/43501638>
- Stasz, C. (1976). Contract Menu Grading. *Teaching Sociology*, 4(1), 49–66.  
<http://10.0.9.3/1317088>
- Staus, N. L., O’Connell, K., & Storksdieck, M. (2021). Addressing the ceiling effect when Assessing STEM out-of-school time experiences. *Frontiers in Education*, 6.  
<https://doi.org/10.3389/educ.2021.690431>
- Stella, J., & Corry, M. (2013). Teaching writing in online distance education: Supporting student success. *Online Journal of Distance Learning Administration*, 16(2), 73-83. <https://hdl.handle.net/1805/21253>
- Stelzner, S. L. (1975). A case for contract grading. *Speech Teacher*, 24(2), 127-132.  
<http://10.0.4.56/03634527509378140>
- Stuckey, M. A., Erdem, E., & Waggoner, Z. (2020). Rebuilding habits: Assessing the impact of a hybrid learning contract in online first-year composition courses. *Journal of Writing Assessment*, 13(2), 1–13.  
<http://journalofwritingassessment.org/article.php?article=159>
- Sulz, L., Gibbons, S. L., Naylor, P.-J., & Wharf-Higgins, J. (2020). Evaluation of an intervention to increase student motivation and enrollment in physical education. *The Physical Educator*, 77(3), 505-530.
- Swaffield, S. (2011). Getting to the heart of authentic assessment for learning. *Assessment in Education: Principles, Policy, and Practice*, 18(4), 433-449.  
<https://doi.org/10.1080/0969594X.2011.582838>
- Swales, J. M. (2016). Reflections on the concept of discourse community. *ASp. La revue du Geras*, 69, 7-19. <https://doi.org/10.4000/asp.4774>
- Taber, K. S. (2018). The use of Cronbach’s alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- Tannock, S. (2017). No grades in higher education now! Revisiting the place of graded assessment in the reimagination of the public university. *Studies in Higher Education*, 42(8), 1345-1357. <https://doi.org/10.1080/03075079.2015.1092131>
- Teo, N. Q. X., Lim, T., & Tong, E. M. W. (2023). The humble estimate: Humility predicts higher self-assessment accuracy. *British Journal of Social Psychology*, 62(1), 561–582. <https://doi.org/10.1111/bjso.12545>
- Tippin, G. K., Lafreniere, K. D., & Page, S. (2012). Student perception of academic

- grading: Persistence, academic orientation, and effort. *Active Learning in Higher Education*, 13(1), 51-61. <https://doi.org/10.1177/1469787411429187>
- Ucar, H., & Kumtepe, A. T. (2019). Effects of the ARCS-V-based motivational strategies on online learners' academic performance, motivation, volition, and course interest. *Journal of Computer Assisted Learning*, 36, 335-349. <https://doi.org/10.1111/jcal.12404>
- van Merriënboer, J. J. G., Jelsma, O., & Paas, F. G. W. C. (1992). Training for reflective expertise: A four-component instructional design model for complex cognitive skills. *Educational Technology Research and Development*, 40(2), 23-43. <https://www.jstor.org/stable/30218588>
- Wah, L. K. (2015). The effects of instruction using the ARCS model and GeoGebra on upper secondary students' motivation and achievement in learning combined transformation. *Asia Pacific Journal of Educators and Education*, 30, 141–158. [http://apjee.usm.my/APJEE\\_30\\_2015/APJEE%2030%20Art%209%20\(141-158\).pdf](http://apjee.usm.my/APJEE_30_2015/APJEE%2030%20Art%209%20(141-158).pdf)
- Wang, Y., Peng, H., Huang, R., Hou, Y., & Wang, J. (2008). Characteristics of distance learners: Research on relationships of learning motivation, learning strategy, self-efficacy, attribution, and learning results. *Open Learning*, 23(1), 17-28. <https://doi.org/10.1080/02680510701815277>
- Wang, S., & Zhang, D. (2020). Perceived teacher feedback and academic performance: The mediating effect of learning engagement and moderating effect of assessment characteristics. *Assessment & Evaluation in Higher Education*, 45, 973-987. <https://doi.org/10.1080/02602938.2020.1718599>
- Ward, E. (2021a). An integrated mixed-methods study of contract grading's impact on adolescents' perceptions of stress in high school English: A pilot study. *Assessing Writing*, 48, Article 100508. <https://doi.org/10.1016/j.asw.2020.100508>
- Ward, E. (2021b). Easing stress: Contract grading's impact on adolescents' perceptions of workload demands, time constraints, and challenge appraisal in high school English. *Assessing Writing*, 48, Article 100526. <https://doi.org/10.1016/j.asw.2021.100526>
- Weigle, S. C. (2009). *Assessing writing*. Cambridge University Press.
- White, C. B., & Fantone, J. C. (2010). Pass-fail grading: Laying the foundation for self-regulated learning. *Advances in Health Sciences Education*, 15(4), 469–477. <https://doi.org/10.1007/s10459-009-9211-1>
- Wighting, M. J. (2011). Measuring sense of community and perceived learning among alternative licensure candidates. *Journal of the National Association for Alternative Certification*, 6(1), 4-12. <http://www.jnaac.org/jnaac/>

- Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, 11(3 & 4), 283-289. <https://doi.org/10.1080/10627197.2006.9652993>
- Wolvin, A. D., & Wolvin, D. R. (1975). Contract grading in technical speech communication. *Speech Teacher*, 24(2), 139-142. <http://10.0.4.56/03634527509378142>
- Wood, J. (2021). A dialogic technology-mediated model of feedback uptake and literacy. *Assessment & Evaluation in Higher Education*, 1–18. <https://doi.org/10.1080/02602938.2020.1852174>
- Yang, Y., Oosterhof, A., & Xia, Y. (2015). Reliability of scores on the summative performance assessments. *Journal of Educational Research*, 108(6), 465–479. <https://doi.org/10.1080/00220671.2014.917255>
- Yarber, W. L. (1974). Retention of knowledge: Grade contract method compared to the traditional grading method. *The Journal of Experimental Education*, 43(1), 92–96. <https://www.jstor.org/stable/20150998>
- Yates, N., Gough, S., & Brazil, V. (2022). Self-assessment: With all its limitations, why are we still measuring and teaching it? Lessons from a scoping review. *Medical Teacher*, 44(11), 1296–1302. <https://doi.org/10.1080/0142159X.2022.2093704>
- Zimmerman, Barry 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>



## APPENDIX A

### Student Assessment Rubric

Outcome A: Students are able to distill a primary argument into a single, compelling statement.	
5	The paper foregrounds a succinct, unambiguous, & focused thesis, that is, a <i>central, controlling claim</i> that is: <ul style="list-style-type: none"><li>• <i>arguable</i> (rather than a fact, a recognized truth, or a matter of personal taste),</li><li>• <i>reasoned</i> (e.g., “E-cigarettes should be regulated <i>because</i> ...), and</li><li>• functions as the main <i>result</i> of the research.</li></ul>
4	The paper foregrounds a thesis that is a <i>central, controlling claim</i> but is a bit less compelling, focused, succinct or unambiguous.
3	The paper contains a thesis but, in meeting the stated purpose of the paper, is too broad, too narrow, or lacks adequate focus.
2	The paper contains elements of a thesis (e.g., a central claim, reasons) but fails to bring together these elements in a statement that most readers would recognize as a “thesis.”
1	The paper lacks any sense of a central claim related to the paper’s stated purpose.
Outcome B: Students give a clear purpose and audience.	

5	The paper establishes a clear, specific purpose in relation to <i>impressive</i> knowledge of pertinent research and, in doing so, establishes a strong sense of audience (viz., the paper demonstrates knowledge of an "academic conversation" and is tailored to take part in that conversation).
4	The paper establishes its purpose in relation to <i>ample</i> knowledge of pertinent research and, in doing so, establishes a clear sense of audience.
3	The paper defines a purpose and establishes a sense of audience based on <i>rudimentary</i> knowledge of pertinent research (viz., the paper demonstrates some awareness that it needs to contribute to an existing academic conversation).
2	The paper maintains a purpose and sense of audience, though not formulated in response to pertinent research (i.e., the purpose is not situated in a conversation).
1	The paper does not exhibit a <i>controlling</i> sense of purpose and audience. The paper exhibits shifts in audience or lacks a clear sense of audience altogether.
Outcome C: Students are able to order major points in a reasonable and convincing manner based on primary argument.	
5	From the beginning, the paper provides readers with a clear sense of direction (organization). The paper maintains that sense of direction by using cues (e.g., transitions) to guide readers from one step to the next. The

	conclusion of the paper carries the sense that the paper's stated purpose has been achieved.
4	The paper provides readers with a clear sense of direction though that sense of direction is not always maintained clearly through the use of discursive cues.
3	The paper contains some but minimal effort to give readers a sense of its direction.
2	The paper seems to have some sense of direction but does nothing to make that direction clear to readers.
1	The paper lacks a sense of direction and, thus, lacks global organization.
Outcome D: Students are able to develop their ideas using appropriate rhetorical patterns (e.g., narration, example, comparison/contrast, classification, cause/effect, definition).	
5	The paper is <i>impressive</i> in its development of arguments, e.g., by defining key words, by clarifying ideas through the use of examples or the use of comparison, by clarification through use of narration or classification.
4	The paper develops several of its arguments, e.g., by defining key words, by clarifying ideas through the use of examples or the use of comparison, by clarification through use of narration or classification.
3	The paper reflects an understanding of the need to develop ideas but develops only one or two.

2	The paper reflects some but inadequate effort at developing its ideas.
1	The paper shows no effort at developing its ideas.
Outcome E: Students are able to manage and coordinate basic information gathered from multiple secondary sources.	
5	<p>The paper makes <i>impressive</i> use of basic information from multiple, reliable sources to</p> <ul style="list-style-type: none"> <li>• make clear the situation, problem, or question that the paper engages;</li> <li>• introduce readers to different positions in an academic “conversation” regarding the situation, problem, or question; and</li> <li>• provide supporting evidence for the paper’s arguments.</li> </ul> <p>All of the information from sources is well integrated and is appropriately attributed to the sources.</p>
4	<p>The paper makes <i>good</i> use of basic information from multiple, reliable sources to</p> <ul style="list-style-type: none"> <li>• make clear the situation, problem, or question that the paper engages;</li> <li>• introduce readers to different positions in an academic “conversation” regarding the situation, problem, or question; and</li> <li>• provide supporting evidence for the paper’s arguments.</li> </ul> <p>Most of the information from sources is well integrated and appropriately attributed to the sources.</p>

3	The paper provides supporting information from multiple sources, but the reliability or appropriateness of some sources would be regarded as questionable by likely readers of the paper. Information from sources is adequately integrated and attributed to the sources.
2	The paper provides supporting information, but only from one source or from multiple unreliable sources. Information is poorly integrated and/or poorly attributed to the sources.
1	The paper fails to use basic information gathered from multiple, reliable sources. Information is not integrated and is not attributed to the sources.
Outcome F: Students are able to employ correct diction, syntax, usage, grammar, and mechanics.	
5	The paper reflects a degree of mastery over diction, grammar, syntax, and usage in formal written English, as well as a degree of mastery over other conventions appropriate to academic papers (e.g., APA or MLA documentation style), including the appropriate mechanics for citing sources.
4	In spite of a few errors, the paper reflects control over diction, grammar, syntax, and usage in formal written English, as well as control of conventions appropriate to the purpose of the paper, including the appropriate mechanics for citing sources.

3	In spite of numerous errors, the paper reflects basic control over formal written English, as well as control of conventions appropriate to the purpose of the paper, including the appropriate mechanics for citing sources.
2	The paper contains an obtrusive number of grammatical, syntactic, or usage, and provides minimal mastery of the mechanics for citing sources.
1	The paper reflects a significant lack of control over formal written English (including diction, grammar, usage, and mechanics).

## APPENDIX B

### Project 1 – Summary & Response

#### Learning Objectives

- Students will understand academic writing as a conversation about topics of consequence.
- Students will understand their responsibilities as writers – to accurately cite the work of other writers, to provide their audience with reliable information, and to consider multiple points of view.
- Students will understand academic writing as governed by the conventions of specific discourse communities.
- Students will become more critical readers, learning strategies for previewing, annotating, summarizing analyzing, and critiquing texts.
- Students will improve their ability to write clear and compelling thesis statements.
- Students will develop the skill of constructive critique, focusing on higher order concerns during peer workshops.
- Students will understand the distinction between revising and editing.

#### Assignment

For Project 1, you will summarize and then respond to one of the readings from Chapter 18 in *They Say / I Say (TSIS)*. In your essay, you will summarize the reading and then respond to it by discussing how your own experiences and knowledge have led you to either agree, disagree, or both agree and disagree with the author. Carefully read the example essays that I have posted under Course Content in D2L, as they will help you to understand the expectations for the assignment.

You may use any of the readings in the chapter (even if they were not assigned).

### **Assigned Readings**

- “Is Google Making Us Stupid?” (pp. 424-440)
- “Smarter Than You Think: How Technology Is Changing Our Minds for the Better” (pp. 441-461)
- "Does Texting Affect Writing" (pp. 462-473)
- "No Need to Call" (pp. 505-524)
- “Go Ahead: Waste Time on the Internet” (pp. 500-504)

### **Unassigned Readings**

- "How I Learned to Love Snapchat" (pp. 474-479)
- "Google, Democracy, and the Truth about Internet Search" (pp. 480-499)
- “Does a Protest’s Size Matter?” (pp. 525-529)

### **Requirements**

- Length: 1,000-1,200 words
- Include 3-4 direct quotes from the reading you chose for this assignment.
- Your thesis should state whether you agree, disagree, or both agree and disagree with the author. You must respond to the author’s argument.
- Provide an adequate summary for your reader, but do not allow the summary to dominate the essay. The bulk of the essay should present your own ideas.
- Properly introduce, present, and cite all direct quotes.
- Include a Works Cited page in which you cite the reading that you chose for this assignment.



- You must adhere to the formatting guidelines set forth in *The MLA Handbook*, 8th edition. Be sure that all margins measure 1 inch and that you use Times New Roman 12-point font. You also should follow MLA formatting guidelines regarding the page heading, running header, page numbering, etc.

### **Evaluation Criteria**

- Is the writer's purpose/position clear?
- Does the writer position him/herself within an existing debate/conversation?
- Does the writer spend too many, too few, or just enough words discussing the argument expressed in the selected reading?
- Does the writer offer sufficient (quantity) and compelling (quality) support for his/her own position?
- Is the essay effectively organized?
- Are the paragraphs adequately developed?
- Is the tone appropriate to the essay's purpose?
- Is there evidence of attention to language, of a conscious attempt to employ rhetorical strategies to achieve a certain effect?
- Does the essay contain errors in grammar, punctuation, spelling, and/or mechanics?
- Does the writer smoothly incorporate source material, using signal phrases and transitions?
- Does the writer accurately cite all sources both in the text of the essay and on the Works Cited page?

## APPENDIX C

### Project 2 – Enter the Conversation

#### Learning Objectives

- Students will understand academic writing as a conversation about topics of consequence.
- Students will understand their responsibilities as writers – to accurately cite the work of other writers, to provide their audience with reliable information, and to consider multiple points of view.
- Students will understand academic writing as governed by the conventions of specific discourse communities.
- Students will become more critical readers, learning strategies for previewing, annotating, summarizing analyzing, and critiquing texts.
- Students will acquire informational literacy – the ability to locate and evaluate source material.
- Students will improve their ability to write clear and compelling thesis statements.
- Students will develop the skill of constructive critique, focusing on higher order concerns during peer workshops.
- Students will understand the distinction between revising and editing.

#### Assignment

For Project 2, you will summarize and then respond to one of the readings from this unit (or the video, *College Inc.*). In your essay, you will summarize the reading/video and then respond to it by discussing how your own experiences and knowledge have led you to either agree, disagree, or both agree and disagree with the author and by including the

opinions of third parties (i.e., by incorporating secondary sources), which is discussed in more detail below.

Most of the readings can be found in your textbook. However, I also assigned a couple of outside readings and the video, *College, Inc.*, which are posted under Course Content in D2L. In addition to the assigned readings (or the video), you may choose any of the other readings from Chapter 17 in *They Say / I Say*. Choose the one that you best understand. Carefully read the example essays that I have posted under Course Content in D2L, as they will help you to understand the expectations for the assignment.

Project 2 is similar to the previous essay, with two additions:

1. Rather than responding to the selected reading/video with your own opinion only, you will add other people's voices to the conversation by including two secondary sources (i.e., in addition to the selected reading/video). You will use quotes both from the selected reading/video and from your two secondary sources to support your assertions.

Your secondary sources can be other readings from this unit. For example, in "Two Years Are Better Than Four," Liz Addison is responding to Rick Perlstein's argument in "What's the Matter with College?" Therefore, you might choose to discuss their opposing views. Instead, you might choose articles you find through one of the library databases, an article in another textbook, a radio show, a podcast, or a video. You are not required to use scholarly sources (i.e., peer-reviewed academic journals from a library database), but your sources should be reliable. For example, you may use *The New York Times* or *Forbes*, but you

cannot use Wikipedia, a blog, or a personal website. If in doubt about your sources, please ask me.

2. You will begin to answer the questions, “So what?” and “Who cares?” by considering your intended audience and by linking your argument to issues that your readers already care about.

### **Assigned Readings/Videos**

- “What’s the Matter with College?” (posted under Content in D2L)
- "Two Years are Better Than Four" (pp. 365-368 in *TSIS*)
- "The New Liberal Arts" (pp. 336-343 in *TSIS*)
- "Blue-Collar Brilliance" (pp. 377-389 in *TSIS*)
- "Are Too Many People Going to College?" (pp. 344-364 in *TSIS*)
- "Why Do You Think They're Called For-Profit Colleges?" (posted under Content in D2L)
- *College, Inc.* (link posted under Content in D2L)

### **Unassigned Readings**

- “Should Everyone Go to College?” (pp. 318-335 in *TSIS*)
- “Hidden Intellectualism” (pp. 369-376 in *TSIS*)
- “Shut Up about Harvard” (pp. 390-397 in *TSIS*)
- “On the Front Lines of a New Culture War” (pp. 398-419 in *TSIS*)

### **Requirements**

- Length: 1,000-1,200 words
- Include 3-4 direct quotes from the reading/video that you are responding to.
- Include at least 1 quote from each of your two secondary sources.

- Your thesis should state whether you agree, disagree, or both agree and disagree with the author. You must respond to the author's argument.
- Provide an adequate summary for your reader, but do not allow the summary to dominate the essay. The bulk of the essay should present your own ideas.
- Properly introduce, present, and cite all direct quotes.
- Include a Works Cited page in which you cite both the reading/video that you are responding to and your two secondary sources.
- You must adhere to the formatting guidelines set forth in *The MLA Handbook*, 8th edition. Be sure that all margins measure 1 inch and that you use Times New Roman 12-point font. You also should follow MLA formatting guidelines regarding the page heading, running header, page numbering, etc.

### **Evaluation Criteria**

- Is the writer's purpose/position clear?
- Does the writer position him/herself within an existing debate/conversation?
- Does the writer answer the questions, "So what?" and "Who cares?"
- Does the writer spend too many, too few, or just enough words discussing the argument expressed in the selected reading?
- Does the writer offer sufficient (quantity) and compelling (quality) support for his/her own position?
- Is the essay effectively organized?
- Are the paragraphs adequately developed?
- Is the tone appropriate to the essay's purpose?

- Is there evidence of attention to language, of a conscious attempt to employ rhetorical strategies to achieve a certain effect?
- Does the essay contain errors in grammar, punctuation, spelling, and/or mechanics?
- Does the writer smoothly incorporate source material, using signal phrases and transitions?
- Does the writer accurately cite all sources both in the text of the essay and on the Works Cited page?

## APPENDIX D

### Project 3 – Annotated Bibliography

#### Learning Objectives

- Students will learn to create citations that conform to The MLA Handbook, 8th edition citation style.
- Students will learn to evaluate internet research sources and distinguish scholarly from popular sources.
- Students will learn to read and understand scholarly articles.
- Students will learn to summarize and paraphrase scholarly articles.
- Students will learn to analyze and respond to scholarly articles.

#### Assignment

The Annotated Bibliography and Essay 3 are the capstones of the course. For the Annotated Bibliography assignment, you will choose a topic related to food and eating addressed in Chapter 20 in *They Say / I Say (TSIS)* that you would like to explore in depth. You may choose an issue from any of the readings in the chapter (even if they were not assigned).

#### Assigned Readings

- “Escape from the Western Diet (pp. 624-631 in *TSIS*)
- “Food as Thought: Resisting the Moralization of Eating” (pp. 641-646 in *TSIS*)
- “Don’t Blame the Eater” (pp. 647-650 in *TSIS*)
- “What You Eat Is Your Business” (pp. 651-655 in *TSIS*)
- “How Junk Food Can End Obesity” (pp. 681-712 in *TSIS*)
- “Why Don’t Convenience Stores Sell Better Food?” (pp. 632-640 in *TSIS*)

## Unassigned Readings

- “The Extraordinary Science of Addictive Junk Food” (pp. 656-680 in *TSIS*)
- “Expanding the National School Lunch Program to Higher Education” (pp. 713-729 in *TSIS*)

The Annotated Bibliography assignment is designed to prepare you for Essay 3, which is a formal research essay with 4 scholarly sources. The Annotated Bibliography provides the research basis for the essay, and you will choose 4 sources from your Annotated Bibliography to use in Essay 3. You will use the same topic for both assignments. Choose your topic wisely because you will be researching and writing about it for the remainder of the semester.

The assignment is not difficult, but it is time-consuming and requires close reading and meticulous documentation of sources. If you work hard and meet the requirements, then you will earn a good grade. If you procrastinate and put forth minimal effort, then you will earn a poor grade. Be sure to budget your time accordingly.

Please read these instructions carefully. This assignment has multiple parts.

1. You will choose one of the food and eating topics addressed in the readings for Unit 3. Again, Essay 3 will be based on the research you gather for the Annotated Bibliography, so the topic you choose for the Annotated Bibliography will be the topic for your final paper.
2. After choosing your topic, you will research it extensively. During this process, you should consider your opinions about your topic and begin to formulate a thesis for your final paper. Remember: The title of this course is Research &



Argumentative Writing, so your final essay must be argumentative. You will not inform the reader about the topic; you will persuade the reader.

3. After completing your research, you will select 8 scholarly articles from peer-reviewed academic journals that address the topic you plan to write about in your final essay.

Important: You are required to use only scholarly articles from peer-reviewed academic journals for this assignment. Failing to use scholarly sources for these assignments will result in a grade of *revise* for the assignment. Therefore, it is crucial that you use scholarly sources.

If you are using Google to locate sources, then it is highly unlikely that you are choosing scholarly articles. Your safest option is to use the library databases (e.g., ProQuest, JStor, and Academic OneFile) to find sources. If you do not know how to access/use the library databases, please see the "Library Research Tutorials" section in Content. If you do not know how to distinguish scholarly journals from popular periodicals, please see the "Evaluating Research Sources" module under Content. In addition, several of the DisQs in this unit will teach you to evaluate sources and distinguish scholarly sources from non-scholarly ones.

4. Next, you will annotate the articles. Each summary should be at least 250 words. Please see the document "What Is an Annotated Bibliography (posted under Course Content) for detailed information about writing annotations.

5. After annotating the articles, you will create citations that conform to The MLA Handbook, 8th edition citation style.
6. Finally, you will compile the annotations and the citations into an Annotated Bibliography and dropbox the assignment by the due date listed in the course schedule.

### **Requirements**

1. You should use the same research topic for both the Annotated Bibliography and for Essay 3.
2. Your Annotated Bibliography must contain 8 entries.
3. You are required to use scholarly articles from peer-reviewed academic journals.
4. You must have both a summary and an evaluation for each of your 8 sources.  
Your summaries must be detailed and thorough. Each annotation summary should be at least 250 words.
5. You must properly cite each source using the MLA Handbook, 8th edition citation style.
6. Be careful when directly quoting sources. Do not overquote. A summary is a summary. It is acceptable to include one or two small quotes if your summary is primarily in your own words. However, long quotes that overwhelm the summary are not acceptable.
7. Your Annotated Bibliography must have an original (i.e., interesting) title.
8. Your summaries must be written in the literary present tense.

## Revisions

As with all assignments in this course, your submissions must meet certain minimum requirements. Failing to meet the requirements of the assignment will require you to revise the assignment. The following is a list of major revision standards for the assignment:

1. You are required to use peer-reviewed articles from academic journals. Using non-scholarly articles will result in a grade of *revise* for the assignment.
2. You are required to include 8 scholarly articles from peer-reviewed academic journals in the Annotated Bibliography. Failing to include the required number of sources will result in a grade of *revise* for the assignment.
3. You are required to make a good-faith effort to use correct MLA citation style. Providing a list of links without making an effort to create MLA citations will result in a grade of *revise* for the assignment.

Note: The Annotated Bibliography is the most challenging assignment this semester, but I think that you will find it to be labor-intensive rather than difficult. If you follow instructions and work hard, you will do well!

## APPENDIX E

### Project 4 – Contribute to the Scholarship

#### Learning Objectives

- Students will understand academic writing as a conversation about topics of consequence.
- Students will understand their responsibilities as writers – to accurately cite the work of other writers, to provide their audience with reliable information, and to consider multiple points of view.
- Students will understand academic writing as governed by the conventions of specific discourse communities.
- Students will become more critical readers, learning strategies for previewing, annotating, summarizing analyzing, and critiquing texts.
- Students will acquire informational literacy – the ability to locate and evaluate source material.
- Students will improve their ability to write clear and compelling thesis statements.
- Students will develop the skill of constructive critique, focusing on higher order concerns during peer workshops.
- Students will understand the distinction between revising and editing.

#### Assignment

For Project 4, you will contribute to the current body of scholarship about the topic you chose for your Annotated Bibliography. In your essay, you will persuade the reader to accept the argument set forth in your thesis using a combination of your own opinions

and the opinions of third parties (i.e., by incorporating secondary sources), which is discussed in more detail below.

### **Requirements**

1. Length: 1,200-1,500 words (which is longer than the previous essays)
2. You must use the topic from your Annotated Bibliography. The purpose of the Annotated Bibliography assignment was to select, summarize, and analyze your sources for this essay.
3. Your thesis must be argumentative (i.e., persuasive).
4. You must include at least four secondary sources from your Annotated Bibliography.
5. You must include at least one quote from each of your four secondary sources.
6. You must use only articles from peer-reviewed scholarly journals for this essay. Failure to use scholarly sources will result in a grade of *revise* for the assignment. Therefore, it is crucial that you use scholarly sources.
7. You must include a Works Cited page that conforms to *The MLA Handbook*, 8th edition citation style.
8. You must include correct MLA parenthetical/in-text citations.
9. You must properly introduce, present, and cite all direct quotes.
10. Your tone should be formal.

### **Evaluation Criteria**

- Is the writer's purpose/position clear?
- Does the writer position him/herself within an existing conversation/debate?
- Does the writer answer the questions, "So what?" and "Who cares?"

- Does the writer spend too many, too few, or just enough words discussing others' views?
- Does the writer offer sufficient (quantity) and compelling (quality) support for his/her own view?
- Is the essay effectively organized?
- Are the paragraphs adequately developed?
- Is the tone appropriate to the essay's purpose?
- Is there evidence of attention to language, of a conscious attempt to employ rhetorical strategies to achieve a certain effect?
- Does the essay contain errors in grammar, punctuation, spelling, and/or mechanics?
- Does the writer smoothly incorporate source material, using signal phrases and transitions?
- Does the writer accurately cite all sources both in the text of the essay and on the Works Cited page?

## APPENDIX F

### Course Interest Survey

Instructions:

There are 34 statements in this questionnaire. Please think about each statement in relation to the class you have just taken and indicate how true it is. Give the answer that truly applies to you, and not what you would like to be true, or what you think others want to hear.

Think about each statement by itself and indicate how true it is. Do not be influenced by your answers to other statements.

Record your responses on the answer sheet that is provided and follow any additional instructions that may be provided in regard to the answer sheet that is being used with this survey.

Use the following values to indicate your response to each item.

1 (or A) = not true

2 (or B) = slightly true

3 (or C) = moderately true

4 (or D) = mostly true

5 (or E) = very true

1. The instructor knows how to make us feel enthusiastic about the subject matter of this course.
2. The things I am learning in this course will be useful to me.
3. I feel confident that I will do well in this course.

4. This class has very little in it that captures my attention.
5. The instructor makes the subject matter of this course seem important.
6. You have to be lucky to get good grades in this course.
7. I have to work too hard to succeed in this course.
8. I do NOT see how the content of this course relates to anything I already know.
9. Whether or not I succeed in this course is up to me.
10. The instructor creates suspense when building up to a point.
11. The subject matter of this course is just too difficult for me.
12. I feel that this course gives me a lot of satisfaction.
13. In this class, I try to set and achieve high standards of excellence.
14. I feel that the grades or other recognition I receive are fair compared to other students.
15. As a student in this class, I am curious about the subject matter.
16. I enjoy working in this course.
17. It is difficult to predict what grade the instructor will give my assignments.
18. I am pleased with the instructor's evaluations of my work compared to how well I think I have done.
19. I feel satisfied with what I am getting from this course.
20. The content of this course relates to my expectations and goals.
21. The instructor does unusual or surprising things that are interesting.
22. The students actively participate in this class.
23. To accomplish my goals, it is important that I do well in this course.
24. The instructor uses an interesting variety of teaching techniques.



25. I do NOT think I will benefit much from this course.
26. I often daydream while in this class.
27. As I am taking this class, I believe that I can succeed if I try hard enough.
28. The personal benefits of this course are clear to me.
29. My curiosity is often stimulated by the questions asked or the problems given on the subject matter in this class.
30. I find the challenge level in this course to be about right: neither too easy nor too hard.
31. I feel rather disappointed with this course.
32. I feel that I get enough recognition of my work in this course by means of grades, comments, or other feedback.
33. The amount of work I have to do is appropriate for this type of course.
34. I get enough feedback to know how well I am doing.

*Note: Reprinted from Keller, J. (2010). Motivational design for learning and performance: the ARCS model approach. Springer.*

## APPENDIX G

### Volition for Learning Scale

Use the following values to indicate your response to each item.

1 (or A) = completely disagree

2 (or B) = somewhat disagree

3 (or C) = neither agree nor disagree

4 (or D) = somewhat agree

5 (or E) = completely agree

1. My commitment to achieve the goals in this class was strong relative to the goals in my other classes.
2. I set up goals for my learning.
3. I was confident that I could avoid obstacles while doing my work.
4. I was prepared to work hard to achieve my goals no matter what my other classes required.
5. I was able to prepare a study plan that listed concrete tasks.
6. I kept my feelings under control while working to complete this class.
7. I added more effort to stay on task if my focus on my goal in this class began to decline.
8. I was able to avoid being distracted by competing goals.
9. I was able to create a setting free of uncontrollable distractions.
10. I was able to know when to stop looking for more information to prepare for an exam.

11. I didn't let social pressure affect my performance.
12. I anticipated personal or social events that might cause me to get behind.
13. When my motivation decreased, I was able to think of things to do to build it back up again.

*Note:* Reprinted from Keller, J. M., Ucar, H., & Kumtepe, A. T. (2019). Development and validation of a scale to measure volition for learning. *Open Praxis, 12*(2), 161–174.  
<https://doi.org/10.5944/openpraxis.12.2.1082>

## APPENDIX H

### CAP Perceived Learning Scale

Directions:

A number of statements that students have used to describe their learning appear below.

Some statements are positively worded, and others are negatively worded. Carefully read each statement and then place an X in the appropriate column to the right of each statement to indicate how much you agree with the statement, where lower numbers reflect less agreement and higher numbers reflect more agreement. There is no right or wrong response to each statement and your course grade will not be influenced by how you respond. Do not spend too much time on any one statement but give the response that seems to best describe the extent of your learning. It is important that you respond to all statements.

Using the scale to the right, please respond to each statement below as it specifically relates to your experience in this course.

Not at all

Very much so

0    1    2    3    4    5    6

1. I can organize course material into a logical structure.
2. I cannot produce a course study guide for future students.
3. I am able to use physical skills learned in this course outside of class.
4. I have changed my attitudes about the course subject matter as a result of this course.
5. I can intelligently critique the texts used in this course.

6. I feel more self-reliant as the result of the content learned in this course.
7. I have not expanded my physical skills as a result of this course.
8. I can demonstrate to others the physical skills learned in this course.
9. I feel that I am a more sophisticated thinker as a result of this course.

*Note:* Reprinted from Rovai, A. P., Wighting, M. J., Baker, J. D., Grooms, L. D.

Development of an instrument to measure perceived cognitive, affective, and psychomotor learning in traditional and virtual classroom higher education settings.

*Internet and Higher Education*, 12, 7-13. <https://doi.org/10.1016/j.iheduc.2008.10.002>

## APPENDIX I

### IRB Approval

IRB #: PRO-FY2022-175

Title: UNGRADING WRITING: EFFECTS ON MOTIVATION, VOLITION, AND PERCEIVED LEARNING

Creation Date: 11-2-2021

End Date:

Status: **Approved**

Principal Investigator: Jennifer Wilson

Review Board: University of Memphis

Sponsor:

---

#### Study History

---

Submission Type	Initial	Review Type	Exempt	Decision	<b>Exempt</b>
Submission Type	Modification	Review Type	Exempt	Decision	<b>Approved</b>
Submission Type	Renewal	Review Type	Unassigned	Decision	

---

#### Key Study Contacts

---

Member	Jennifer Wilson	Role	Principal Investigator	Contact	<a href="mailto:jbwilson5@memphis.edu">jbwilson5@memphis.edu</a>
Member	Jennifer Wilson	Role	Primary Contact	Contact	<a href="mailto:jbwilson5@memphis.edu">jbwilson5@memphis.edu</a>
Member	Andrew Tawfik	Role	Co-Principal Investigator	Contact	<a href="mailto:aatawfik@memphis.edu">aatawfik@memphis.edu</a>

---

## APPENDIX J

### Recruitment Announcement

## Participants Needed!

In this course, you will have the opportunity to participate in a study concerning the effects of contract grading in online freshman composition courses. Please consider participating! The details of the study are provided below.

**Study Description and Purpose:** You are being asked to participate in this research study because I am evaluating the effects of contract grading on motivation, volition, and perceived learning in online freshman composition courses. If you choose to participate in the study, you will be asked to respond to three questionnaires consisting of statements related to your opinions about the course content, materials, and assignments. The questionnaires will be administered three times, once during the first week of the semester, once during the fourth week of the course, and once during the eighth week of the semester. Finally, you will be asked to provide simple demographic information about yourself.

**Target Participant Pool:** Adults 18 and over

**Risks & Discomforts:** You should not feel any more discomfort from this study than you would in your everyday life.

**Benefits:** Though you may not personally benefit from this study, future students have the potential to benefit from the results of this study.

**Additional Information:** Aside from giving up a few hours of your time, there are no other expected costs to you. Once you have completed the questionnaires, your responses will be sent to the researchers.

**Compensation:** You will be entered into a drawing for one of seven \$25 Amazon gift cards. You also will earn credit in the course for your participation.

Once you have completed the questionnaires, you will take a screenshot of the confirmation page. Submit the screenshots to you instructor by uploading them to the appropriate dropboxes in D2L.

Please enter the survey by clicking the appropriate links in D2L. You will be given a chance to read the informed consent to assist you in making a final determination to participate.

Sincerely,

Professor Wilson

**Primary Investigator:** Jennifer Wilson

**PI Department & College:** Instructional Design and Technology Program,  
College of Education, University of Memphis

**Protocol Title:** Ungrading Writing: Effects on Motivation, Volition, and Perceived Learning

**Protocol ID:** PRO-FY2022-175

**Approval Date:** 11/16/2021

**Contact Information:** [jbwilson5@memphis.edu](mailto:jbwilson5@memphis.edu)



## APPENDIX K

### Ungrading Handbook

The purpose of this handbook is to provide guidance for instructors using labor-based ungrading in online, undergraduate research and argumentative writing courses at the university where the proposed study will take place. Assignments will be evaluated according to the following scale:

- Accept – The assignment meets minimum criteria for credit.
- Revise – The assignment does not meet minimum criteria and must be revised to earn credit.

#### Discussion Posts

Students will complete 25 discussion posts during the semester. In order to earn an *accept* designation, the assignment must adhere to the following criteria:

- The post must meet the minimum word length requirement (200 words).
- The post must address the assigned reading(s) and topic.
- The student must respond to at least two other students.
- The post must be original (i.e., not plagiarized).

#### Peer Reviews

Students will complete four peer reviews during the semester. In order to earn an *accept* designation, the assignment must meet the following criteria:

- The peer review must be complete (i.e., all questions answered).
- Answers to the questions must be in sentence form (i.e., not just a few words).
- The peer review must provide substantive feedback and represent a good faith effort to assist the peer review partner.

### **Introduction Drafts**

Students will complete three introduction drafts during the semester. In order to earn an *accept* designation, the assignment must meet the following criteria:

- The introduction draft must address the assigned reading(s) and topic.
- The introduction draft must be at least ½ page in length.
- The introduction must be original (i.e., not plagiarized).

### **Thesis Statement Workshop**

Students will complete one thesis statement workshop during the semester. In order to earn an *accept* designation, the assignment must meet the following criteria:

- The thesis statement worksheet must be fully completed.
- The thesis statement must address the assigned reading(s) and topic.
- The thesis statement must be argumentative.
- The thesis statement must be original (i.e., not plagiarized).

### **Topic Proposal**

Students will complete one topic proposal during the semester. In order to earn an *accept* designation, the assignment must meet the following criteria:

- The topic proposal worksheet must be fully completed.
- The topic proposal must address the assigned reading(s) and topic.
- The topic proposal must be original (i.e., not plagiarized).

### **Research Proposal**

Students will complete one research proposal during the semester. In order to earn an *accept* designation, the assignment must meet the following criteria:

- The research proposal worksheet must be fully completed.

- The research proposal must address the assigned reading(s) and topic.
- The research proposal must be original (i.e., not plagiarized).

## Essays

Students will complete three essays during the semester. In order to earn a *accept* designation, the assignment must meet the following criteria:

- The essay must address the assigned reading(s) and topic.
- The essay must be 1,000-2,000 words in length.
- The essay must be argumentative.
- The essay must contain a clear thesis statement.
- Each body paragraph must contain a clear topic sentence.
- The essay must include quotes from the course reading(s) selected for the assignment.
- The essay must be effectively organized according to academic conventions (i.e., introduction, multiple body paragraphs, and conclusion).
- Each paragraph must be adequately developed (approximately  $\frac{1}{2}$  to  $\frac{2}{3}$  page).
- The tone and word choice must be appropriate for an academic audience.
- The essay must exhibit correct grammar, spelling, and punctuation. The essay must contain no more than 10 grammatical/mechanical errors.
- Quoted/paraphrased/summarized material must be correctly incorporated into the text and cited according to MLA style.
- The essay must include a Works Cited page.
- The essay must be properly formatted.
- The essay must be original (i.e., not plagiarized).

## **Annotated Bibliography**

Students will complete one annotated bibliography during the semester. In order to earn a *accept* designation, the assignment must meet the following criteria:

- The bibliography must address the assigned reading(s) and topic.
- The bibliography must contain 8 bibliographic entries.
- All sources must be scholarly articles from peer-reviewed academic journals.
- Each bibliographic entry must contain an annotation.
- Each annotation must be at least 250 words in length.
- Quoted/paraphrased/summarized material must be correctly incorporated into the text and cited according to MLA style.
- All bibliographic entries must reflect a good faith effort to create correct citations according to MLA citation style.
- The bibliography must be properly formatted according to MLA style.
- The bibliography must be original (i.e., not plagiarized).

## APPENDIX L

### Essay Rubric

<p>Note: If the top sentence in the box is highlighted, then you performed satisfactorily in that area. If one of the other sentences is highlighted, then this is a problem area for you. Please be sure to scroll to the comments section at the bottom of this rubric, where I will leave detailed feedback.</p> <p style="text-align: center;"><b>Requirements/Formatting</b></p> <p>Essay meets the requirements of the assignment.</p> <p>Essay doesn't meet the length requirement.</p> <p>Your page heading is missing/incorrect.</p> <p>Essay needs a running header (i.e., your last name and the page number ½" from the top right corner)</p> <p>Essay is improperly formatted.</p> <p>Essay should be double-spaced.</p> <p>Essay doesn't address the assigned topic.</p> <p>Essay uses incorrect font. Use Times New Roman 12 pt.</p> <p>The essay needs an original, effective title.</p> <p>Essay doesn't contain the correct type/number of sources.</p>	<p style="text-align: center;"><b>Coherence</b></p> <p>Your coherence is satisfactory.</p> <p>Your introduction needs more development.</p> <p>Your conclusion needs more development.</p> <p>You have included information in the introduction or conclusion that should be in the body of the essay.</p> <p>Some paragraphs lack topic sentences.</p> <p>Consider reordering your paragraphs so that the essay flows better.</p> <p>You need transitions between paragraphs.</p> <p>You need transitions within your paragraphs.</p> <p>You need "wrap-up" sentences that conclude each paragraph.</p> <p>Your tone and word choice are too informal for an academic essay.</p>
<p style="text-align: center;"><b>Focus</b></p> <p>Your focus is satisfactory.</p> <p>Your thesis is unclear.</p> <p>Your thesis is too broad or too narrow.</p> <p>Your thesis has more than one focus (i.e., addresses more than one topic).</p> <p>Your thesis is not an assertion.</p> <p>Your essay does not remain focused on your thesis.</p> <p>Your thesis is not stated in the introductory paragraph (i.e., thesis first appears in one of the body paragraphs or in the conclusion).</p>	<p style="text-align: center;"><b>Research &amp; Documentation</b></p> <p>Your research and documentation are satisfactory.</p> <p>You do not cite the required number of sources.</p> <p>Your sources are not scholarly (i.e., from peer-reviewed academic journals)</p> <p>Your parenthetical (in-text) citations need work.</p> <p>Your Works Cited page needs work.</p> <p>Your paraphrasing/summarizing needs work.</p>

	<p>Your quotations (including block quotes) are improperly formatted.</p> <p>You have “dropped quotes” (i.e., you do not introduce cited material with a signal phrase).</p> <p>Your Works Cited does not start on a new page.</p> <p>Your essay is “research-heavy” (i.e., you allow sources to drown out your own voice).</p> <p>Your essay is plagiarized. Contact me immediately.</p>
<p style="text-align: center;"><b>Organization</b></p> <p>Your organization is satisfactory.</p> <p>The essay is poorly organized.</p> <p>The essay has basic organization but needs to be subdivided into smaller, more specific topics.</p> <p>Some paragraphs overlap and repeat each other.</p> <p>Some paragraphs lack unity (i.e., cover more than one topic).</p> <p>Your organization doesn’t fit the assignment.</p> <p>Lengthen short paragraphs or combine them to make longer ones.</p> <p>Shorten long paragraphs or divide them to create shorter paragraphs.</p>	<p style="text-align: center;"><b>Mechanics</b></p> <p>Your mechanics are satisfactory.</p> <p>You have some careless errors; proofread more carefully.</p> <p>You have problems with subject-verb agreement.</p> <p>You have problems with tense shifts.</p> <p>You have problems with verb forms.</p> <p>You have problems with passive voice.</p> <p>You have problems with run-ons.</p> <p>You have problems with comma splices.</p> <p>You have problems with fragments.</p> <p>You leave out commas after subordinate clauses and prepositional phrases.</p> <p>You leave out commas before coordinating conjunctions.</p>

<p style="text-align: center;"><b>Development</b></p> <p>Your development is satisfactory.  Some paragraphs lack full support.  Your support is not specific enough.  You have left out important points.  You don't fully explain your support.  Your support is not convincing.</p>	<p>You put commas where they are unnecessary.  You are having problems with parallel construction.  You have problems with apostrophes.  You have problems with semicolons.  You have problems with quotation marks.  You have problems with pronoun-antecedent agreement.  You have problems with pronoun reference.  You have problems with dangling modifiers.  You have problems with capitalization.  You have problems with spelling.  You have problems with slang.  You have problems with clichéd phrasing.  You have problems with awkward phrasing &amp; wordiness.  Avoid contractions (e.g., isn't, doesn't, can't).  Avoid second person pronouns (you, your).  Mechanical errors make this essay difficult to read; mechanics are affecting content.</p>
<p><b>Comments:</b></p> <p>Grade: <i>Accept or Revise</i></p>	

## APPENDIX M

### Annotated Bibliography Rubric

General Requirements	Yes/No
Does the annotated bibliography address the assigned topic (i.e., food and eating)?	
Does the annotated bibliography contain 8 bibliographic entries?	
Are all 8 of your sources scholarly?	
Are the bibliographic entries correctly alphabetized (i.e., by author's last name)?	
Do you have an original and interesting title?  Note: The title should grab your reader's attention. Don't use Annotated Bibliography as your title. Be creative!	

MLA Citation Style	Yes/No
Did you make a good faith effort to create a complete MLA citation for each source?  Note: I did not mark every citation error. However, you can check your work by referring to the example citations below. Your citations should <u>exactly</u> conform to the following format:  Chauvenet, Christina and Lindsey S. Tallie. "A Call to Oppose the Child Nutrition and Education Act of 2016 (H.R. 5003)." <i>American Journal of Public Health</i> , vol. 106, no. 12, 2016, pp. 2129-2130. <i>EBSCOhost</i> , doi:10.2105/AJPH.2016.303479. Accessed 4 Apr. 2019.	



<p>Leeman, Jennifer, et al. "Policy, Systems, and Environmental Approaches to Obesity Prevention: Translating and Disseminating Evidence from Practice." <i>Public Health Reports</i>, vol. 130, no. 6, 2015, pp. 616–22. <i>JSTOR</i>, <a href="http://www.jstor.org/stable/43776228">www.jstor.org/stable/43776228</a>. Accessed 4 Apr. 2019.</p> <p>Mirowsky, John, and Catherine E. Ross. "Education, Health, and the Default American Lifestyle." <i>Journal of Health and Social Behavior</i>, vol. 56, no. 3, 2015, pp. 297–306. <i>JSTOR</i>, <a href="http://www.jstor.org/stable/44001141">www.jstor.org/stable/44001141</a>. Accessed 10 Apr. 2019.</p>	
Is the first line of each citation even with the left margin?	
Are the second and subsequent lines of each citation indented by ½ inch?	
<p>Have you correctly cited multi-author works?</p> <p>If there are two authors: Wilson, Jennifer, and Bill Gates.</p> <p>If there are three authors: Wilson, Jennifer, Bill Gates, and Steve Jobs.</p> <p>If there are three or more authors: Wilson, Jennifer, et al.</p>	
Are journals titles and database names italicized?	
Are article titles in quotation marks?	
Did you include the page numbers in each citation?	
<p>Are the volume and issue number provided and correctly formatted (e.g., vol. 6, no. 10)?</p> <p>Note: Be sure that the "v" in "vol." and the "n" in "no." are lowercase. Also, be sure that you used "no." rather than "issue."</p>	
Did you include the library database name?	
Did you include the date of access?	

Is the DOI or the URL included for each bibliographic entry?	
Did you remove all hyperlinks from your citations?	
Are your citations correctly punctuated and capitalized?	

Annotations	Yes/No
Did you include an annotation for all 8 sources?	
Is each annotation (summary and evaluation combined) at least 250 words?	
Does each annotation have both a summary and evaluation of the source?	
Is each annotation well-written and free of major grammar/punctuation errors?	
Do you refer to the authors you cite by last name after you've introduced him/her using his/her full name in the introduction?	
Is the first line of each annotation indented by ½ inch?	
Are the second and subsequent lines of each annotation even with the left margin?	
Are your quotes written exactly as the author intended (word-for-word)?	
Are all quoted passages in quotation marks?	
Are paraphrases and summaries written without changing the meaning of the original text?	
Have you used paraphrasing/summarizing rather than quoting whenever possible? Are all of your quotes the appropriate length (i.e., you haven't "over-quoted")?	
When discussing your secondary sources, do you use the literary present tense?	

