Journal of the Scholarship of Teaching and Learning, Vol. 25, No. 2, June 2025, pp.17-33.
doi: 10.14434/josotl.v25i2.36399

Accessing Assessment: The Correspondence of Longitudinal Assessment Tools to Grades in a Writing-Intensive Course

Christopher R. Roland
University of Central Arkansas

Jessica R. Snitko
Embry-Riddle Aeronautical University

Joshua Boyd
Purdue University

Abstract: Researchers have long cast grades as an unreliable proxy for student learning despite their enduring use in higher education. Grounding assessment instruments in student learning outcomes (SLOs) provides an accurate and accessible means of evaluating student performance and guiding instructional practice. Using a SLO-guided pretest/posttest assessment instrument, this study evaluated the correspondence between student learning and grades in a writing-intensive, large-lecture communication course. Learning scores were found to be significantly positively associated with final grades and scores on individual assignments. The study further explored differences in learning and grades between online and in-person course formats. Implications for assessment instrument design are discussed.

Keywords: Assessment, Writing courses, General Education

The evaluation of student learning has served as a central metric of student, instructor, and institutional accountability in higher education. Letter and percentage-based grading systems have long served as the primary if not exclusive means through which universities appraise academic achievement and alter instructional practice (Schinske & Tanner, 2014). Though they are used by nearly every institution of higher education, grades are nevertheless criticized as inadequate measures (Kleinman et al., 2018; McMorran et al., 2017; Sadler, 2013). Detractors have long cast a pessimistic portrayal of grading systems, arguing that grades are fundamentally unassociated with student learning (Arum & Roksa, 2011). This bleak account of grades has received its share of criticism, citing measurement issues and the scope of the critique (Lui et al., 2012). Course evaluations often prompt comments that students have learned a lot, even if their grades do not seem to reflect that, and students sometimes also complain that they have earned good grades but do not necessarily feel confident in their learning.

The purpose of this study is to affirm the presumed but largely unsubstantiated relationship between grading systems and student learning. This study outlines a procedure for building a pretest/posttest assessment instrument of student learning for a large-lecture, writing-intensive communication course. Through grounding the performance-based, low-stakes assessment instrument in specific course learning outcomes, we argue that it is possible to reliably determine the extent to which scores of learning achievement correspond to grades. Additionally, this study explores how the assessment instrument and grades differ between online and in-person instruction for the same course. This paper first describes the limitations of grades as a representative for student learning. Next, it reviews the principles of assessing learning in writing-intensive courses. Finally, it reports on a study that evaluates how a pretest/posttest instrument grounded in learning outcomes can inform instructional practice.

Grade Motivation and Learning

Course grades have long served as the most common proxy for student learning achievement. This is based on the premise that aggregated exams, essays, and reports provide objective and reliable assessment for student learning (Schinske & Tanner, 2014). The need for consistent evaluation facilitated the widespread adoption of point and percentage-based grading systems among 19th century universities, even though early criticisms acknowledged the subjectivity and inconsistency of this system (Cureton, 1971; Schneider & Hutt, 2014). In contemporary higher education, the use of letter grades has developed as the predominant metric of student evaluation (Schneider & Hutt, 2014). Kleinman and colleagues (2018) studied 620 universities in the U.S. and found that 98% used letter grade systems.

Despite the common practice and assumed utility of course grades, their efficacy as evaluative tools have been widely criticized. Detractors have long questioned the ability of course grades to reflect student learning accurately (McMorran et al., 2017; Schinskey & Tanner, 2014). Common criticisms have alleged that grades are unreliable and fail to be replicated across instructors (Sadler, 2013). Similarly, scholars have demonstrated wide-spread grade inflation for nearly as long as letter grades have been adopted, which undermines the consistency that these systems purport to produce (Jewell et al., 2013; Rojstaczer & Healy 2012). While letter grades emerged due to a motivation for inter-institutional consistency, multiple scholars have suggested that these systems fail to provide an objective appraisal of student learning.

Student motivations to pursue grades rather than intrinsic learning may compound the challenges of treating grades as an accurate assessment of learning. Job recruiters have acknowledged college GPA as a primary factor influencing job placement (Cole et al., 2003; Imose & Barber, 2015). Student confidence in job placement is likewise positively associated with GPA (Qenani et al., 2014). Accordingly, students have an extrinsic motivation to pursue grades rather than having an intrinsic desire for learning (Dahlgren et al. 2009; Tippin et al., 2012). Grade motivation can promote surface-level learning among students who may simply memorize or replicate course content (Lindblom-Ylänne et al., 2019; Spada & Moneta, 2012). The consequence of grade motivation is an undermining of the deeper comprehension and application of course concepts which is antithetical to the common criteria of learning adopted by most universities (Campbell & Cabrera, 2014). Consequently, students may prioritize non-cognitive skills like effort (Kelly, 2008) or extra credit (Pynes, 2014) which inflate grades but may not reflect learning. The drive for students to pursue grades above learning results in a nebulous relationship between contemporary grading systems and student learning.

Instructors may be similarly motivated to award high grades not necessarily connected to student learning, which further complicates assessment. One factor influencing instructor orientation toward grades is the emergence of self-report course evaluations (Student Evaluations of Teaching or SETs) alongside modern grading systems (see Uttl et al., 2017). Despite myriad criticisms, SETs have been widely and persistently used as a central evaluative criterion for faculty hiring, promotion, and firing decisions (Spooren et al., 2013). Given the weight ascribed to SETs, instructors are incentivized to design classes that are lenient on grades to bolster course evaluations (Stroebe, 2016). Recent evidence suggests that SET scores are positively associated with courses that are lenient, require minimal effort, and promise high grades (see Stroebe, 2020 for review). As instructors are compelled to be cognizant of SETs, this could disincentivize scrutiny into the extent to which grades and student learning correspond.

Association of Course Grades and Student Learning

A common critique of course grades as an evaluative tool is concern over their efficacy to measure the achievement of learning objectives (Campbell & Cabrera, 2014). In recent years, assessing student learning has received increased attention, but critics question the validity of assigning grades for the purpose of measuring course outcomes (Clayson, 2009; Sadler, 2013; Schinske & Tanner, 2014). In their book *Academically Adrift,* Arum and Roksa (2011) concluded that, while hours spent on coursework and learning gains have decreased, college grades and graduation rates are increasing. Another study explored how the three NSSE (National Survey of Student Engagement) Deep Approaches to Learning (DAL) contribute to academic rigor at the collegiate level (Campbell & Cabrera, 2014). Specifically, the researchers wanted to determine whether DAL constructs are related to students' GPA. Results demonstrated that Deep Learning did not predict student GPA, indicating that student learning may not have the presumed association with grades.

Despite research suggesting that grades are not useful tools for assessing student learning, they continue to be widely used (McMorran et al., 2017). One explanation for the over-reliance on grades is that they have been socially constructed as a necessary component to measure student performance and are considered a tool to hold instructors accountable across universities (Lynch & Hennessy, 2017). But as the ideal measure of educational success has transitioned from an emphasis on quality of instruction to an emphasis on quality of learning, there is increasing pressure for faculty to provide evidence that students are gaining the expected knowledge and skills (Swarat et al., 2017).

Whereas grades are used to measure individual student's learning in each class, independent instruments are typically designed to assess whether course instruction is meeting university-wide SLOs (Nusche, 2008). Canfield et al. (2015) tested the reliability and validity of course grades as indicators of the achievement of university-wide general education learning objectives. Results demonstrated that, although grading methods varied across classrooms, grades were effective in evaluating student learning outcomes. The authors argued that their results address the perceived limitations of grades as evaluative tools and provide support for the continued use of grades to measure student learning. However, written assignments such as essay exams and term papers were not included in their dataset. Therefore, the relationship between grades and learning in courses with heavy writing components remains unclear.

Challenges of Assessing Writing-Intensive Courses

Despite the significant role of writing in a college students' education, the relationship between grades and learning in writing courses has not been thoroughly explored (Zheng & Yu, 2019). The act of writing improves students' critical thinking skills such that students become better thinkers throughout the writing process (Bazerman et al., 2005; Boyd, 2010). Furthermore, assessing in-depth writing assignments, as opposed to other types of assignments, allows researchers to evaluate deeper levels of learning (Campbell & Cabrera, 2015) and determine whether students grasp the complexity of the subject (Swarat et al., 2017). Finally, many universities that are faced with accreditation requirements are implementing campus-wide writing programs to improve students' writing skills (Good et al., 2012). Methods for assessing writing-intensive courses must be established to determine whether these programs are having the desired effect.

There is some variation in what constitutes a "writing-intensive course." According to Good and colleagues (2012), a writing-intensive course must include four writing assignments with the completion of at least one assignment requiring instructor feedback throughout the writing and editing process. Farris and Smith (1992) required more stringent guidelines. They argued that a writing-intensive course should meet the following criteria: students write at least 5000 words during

the semester (at least 2000 of which are graded); students receive feedback on assignments with opportunities for revision; at least 50% of the course grade comes from writing assignments.

Current tools that are designed to assess learning are typically a combination of objective tests and student self-reported surveys (Campbell & Cabrera, 2014). Because of these standards, the challenges associated with writing assessment have become even more complex in the 21st century (Zheng & Yu, 2019). One major obstacle is defining writing ability. Slomp (2012) writes, "Like intelligence, the construct of writing ability is broad, multifaceted, situated, contextual, and resistant to a monolithic, stable definition" (p. 81). When assessing writing assignments, not only is it necessary to define "good writing," but scholars must also develop a reliable method by which to measure it (Zheng & Yu, 2019). Furthermore, because expectations for writing vary across fields, it is necessary to develop a method that can be applied across contexts (Haswell, 2000). In their meta-analysis of writing assessment, Zheng and Yu (2019) found that only 5% of the 219 sourced studies focused on developing a new writing assessment method. Therefore, a goal of this study is to apply a previously tested method for writing assessment (Welch et al., 2022) to a new dataset and determine whether there is relationship between students' grades and their achievement of learning objectives in a writing-intensive course.

Grounding Assessment Instruments in Student Learning Outcomes

It has been three decades since the introduction of SLO assessment and the relationship between accreditation and assessment, and yet assessment is rarely a component of institutions' daily teaching practices (Swarat et al., 2017). If the graded assignments in a course are grounded in SLOs, it should result in assessment instruments that are predictive of student grades (Schneider & Preckel, 2017). The assignments in the course assessed in this study are adapted to the specific learning goals of each of the course's units (Boyd, 2010). Furthermore, a five-point holistic grading rubric is used for all assignments. Because assignments are evaluated holistically according to course learning objectives, students' grade in the course should correlate with their learning.

Delucchi (2014) has observed that grades or other classroom performances serve as proxies for learning, but that these proxies are imperfect. To address this problem, his study relied instead on a pretest/posttest design that more accurately measured student learning in an undergraduate statistics class from the beginning to the end of the semester, an approach also taken by Boyd and Morgan (2018) rather than relying on grades or final exams to assess whether students achieved course learning outcomes.

There are several advantages to using a pretest/posttest design rather than relying on grades or self-evaluations. One is a longitudinal advantage–by administering the same assessment instrument at the beginning and end of a course, it is possible to measure changes in learning relative to course learning outcomes. Thus, we define learning scores as the difference between posttest and pretest performance with a higher score indicating greater learning achievement throughout the semester. The other is an absolute value of learning: by the end of the course, a posttest score reveals the extent to which students have mastered course learning outcomes, regardless of how much skill or knowledge students brought with them to the class.

To build the pretest and posttest, we began with the end in mind–the learning outcomes we wanted students to achieve (see Appendix 1). Then we considered what evidence students could provide that they had met those outcomes. This process was completely divorced from testing considerations; the goal was to develop a measure of evidence that was not measured in degrees, but rather was a binary "achieved" or "not achieved" result for each SLO (see Appendix 2). For some learning outcomes, students could answer a simple question to demonstrate achievement or lack of achievement. To support the SLO "Distinguish among qualitative methods of communication

research," for example, students were asked three questions that required them to correctly name or define one of the qualitative methods taught in the class.

To support the writing-related SLOs, a single question could not assess achievement or lack of achievement. For the SLO "Write and support an argument about a communication text," students were shown a short video and given this prompt: "In one well-developed paragraph, write an argument below about this short video." To assess the achievement of this SLO, two independent coders evaluated the degree of achievement of the following elements of writing an argument about a text: (1) Does the paragraph have a clearly stated thesis? (2) Is the thesis about the text, not just the topic? (3) Does the paragraph present an argument rather than just describing the text? A total of 20 points evaluated the achievement of the 6 learning outcomes. The only difference between pretest and posttest was that students viewed a different short video before answering the video prompt. Based on this pretest/posttest format, the following hypotheses were advanced:

> H1 Posttest scores will be positively associated with student (a) final grades, (b) midpoint grades, and (c) performance on major assignments.
> H2 Learning scores will be positively associated with student (a) final grades, (b) midpoint grades, and (c) performance on major assignments.

Assessment across Course Format

The challenges associated with assessment of student learning are further compounded by the proliferation of online instruction. An emerging body of empirical research has investigated the degree to which education quality and learning extends to online instruction, but these findings have, at best, provided mixed results (see Kebritchi et al., 2017 for review). A variety of issues have emerged concerning the time required to develop courses, implications for pedagogical practice, and added challenges for assessment due to the lack of face-to-face contact (Ko & Rossen, 2017; Limperos, et al., 2015; Oncu & Cakir, 2011). For online courses, students experience lower participation (Ching & Hsu, 2015), while instructors often fail to adapt instructional strategies (Chametzky, 2014). Moreover, the novelty of online instruction renders most institutions, professors, and students ill-equipped to evaluate how to improve online courses (Anderson et al., 2011). Recent research has demonstrated effective strategies for design and assessment of online courses (Martin et al., 2019).

The increased adoption of online instruction and mixed findings associated with its capacity to achieve student learning invites further exploration into strategies for assessing online instruction. There is value in directly comparing learning achievement between in-person and online instructional formats. Due to the myriad different student, instructor, and content factors, it stands to reason that an evaluation which can generalize across these factors is warranted. Accordingly, a supporting purpose of the present study was to explore possible differences in learning achievement between online and in-person instruction. Comparing pretest and posttest scores between an online and in-person section would provide some evidence about the relationship between instructional format and learning. This would further affirm that the instrument generalizes across the format of class instruction and thus, the following research question was proposed:

> RQ1 Do online and in-person formats differ in (a) student learning gains, (b) posttest scores, or (c) student performance on major assignments?

While there is value in instructor-based assessment of learning and grades, there is likewise value in exploring student self-appraisal of their performance. Some institutions have moved toward

students engaging in active reflection of their own performance in place or supplement to instructor evaluation via grades (Jaschik, 2011). For some time, researchers have acknowledged the influence of satisfaction on educational outcomes (de Kleijn, et al., 2012). Additionally, some research has acknowledged the positive association between confidence and performance on student grades (Nicholson et al., 2013). Accordingly, this study posed a second research question concerning whether there is a relationship between student self-reported indices of performance (satisfaction and confidence) and observed performance on grades:

> RQ2 Are student opinions about a. course satisfaction and b. confidence associated with their observed class performance?

Course Context for SLO-grounded Assignment Design

In the course studied in this project, all assignments and exam questions are mapped to SLOs (Appendix 1). That is, if an assignment or exam question did not directly support the achievement of a course learning outcome, it was removed from previous course learning activities. For the two major papers, learning outcomes are named in the assignment instructions given to students, so that students know exactly what they are supposed to gain from the assignments. Because the class is part of the university core curriculum in written communication, the first three SLOs (focused primarily on writing) are the most important in the class. The first major paper directly addresses SLOs 1 and 2, and the second major paper directly addresses SLOs 1, 2, and 3. Individual writing assignments and exam questions engage the other 3 SLOs.

In the history of the class, several assignments have been replaced because the result of the assignments did not seem to achieve any SLOs. Sometimes a writing assignment that is interesting or engaging does not actually support an SLO; instructors must sometimes make the hard choice to replace a beloved assignment with something new that more directly contributes to SLOs.

Method

Participants and Design

The sample consisted of a total of 68 students in the same class (same lectures, assignments, and instructional staff) delivered in two different modalities: most of the students were registered for a 16-week face-to-face section (n=50), but the rest were registered for a 16-week online section (n=18). Lectures happened in-person for the larger group but were also live streamed and available online afterwards. All assignments were assigned and collected online. Across sections, an online instrument assessing all course learning outcomes was administered at Timepoint 1 (T1) during the first week of the course and at Timepoint 2 (T2) at week 15. Students received course credit worth 1% of the course grade to incentivize participation. Additionally, a separate instrument assessing self-report satisfaction and confidence was administered during week 16 of the semester.

Measures

Student Learning

Student learning was defined as the aggregate increase in learning achievement by computing difference scores from the end of the semester (T2) and the start of the semester (T1). These outcomes were grounded in six course learning outcomes described in the syllabus (see Appendix 1). Outcomes were assessed using an open-ended 20-point instrument with 3-4 points corresponding to

each outcome (see Appendix 2). The outcomes targeted (1) creating a thesis statement and (2) proposing evidence in support. Second, students (3) identified foundational rhetorical theorists and concepts. Additionally, students (4) distinguished between qualitative research methods and (5) described approaches to critical inquiry. Finally, students (6) applied two methods to analyze a research inquiry in context. Student responses were broken down into relevant points and coded with a dichotomous evaluation of the presence or absence of learning by independent coders. The points were computed to create a global index of learning by aggregating the items in Timepoint 2 and subtracting them from aggregated scores of Timepoint 1.

Course Grades

The class included 15 writing assignments (worth 5 points each) and 15 low-stakes quizzes (repeatable and worth 1 point each), but the highest-value assignments in the class were two exams (25 points each) and two papers (30 points each). Grades were computed at the midpoint (100 points possible) and end of the semester (200 points possible).

Exams. Both exams were timed and involved 50 multiple-choice application-oriented questions. The first exam covered the course's first two units; the final exam was comprehensive, with about ⅓ of the questions from the first half of the course and ⅔ of the questions from the second half of the course.

Term Papers. The first major paper targeted the course's first two learning outcomes. Students chose an unconventional text (i.e., not a speech, editorial, or article) from the previous three months and construct a critical argument about that text, supporting the argument with different parts of the text. Students most frequently chose to write about movies, music videos, and songs. The second major paper addressed the course's first three learning outcomes. Students were presented with a selection of significant speeches from American history and asked to build a critical argument about the speech of their choice. In addition to using support from within the speech, students also incorporated one classical rhetorical concept and one contemporary rhetorical concept into their argument. Scores for major papers were aggregated to create a comprehensive score for papers.

Satisfaction

Students reported their perception that the course was a meaningful learning experience, via a measure of course satisfaction. Satisfaction was defined as the overall subjective appraisal of learning achievement and engagement of course content. Adapting a satisfaction instrument from Gunawardena and Zittle (1997), this was measured using a 10-item, 5-point Likert-type scale with anchors ranging from 1 (strongly disagree) to 5 (strongly agree). Sample items included "As a result of my experience in this course, I would like to participate in a similar course in the future," "I was stimulated to do additional research on topics discussed in class," and "I was able to learn from online discussions." Items were averaged to create a general index of student satisfaction (a = .84).

Confidence

The study included a self-reported appraisal of the extent to which students perceived achievement of learning outcomes. Students were asked to rate the extent to which they perceived gains in the 6 learning outcomes emphasized in the course. This was measured using a 6-item, 5-point Likert-type scale which ranged from 1 (very low confidence) to 5 (very high confidence). Items included

"generating and supporting an argument about a communication text," "composing a clear thesis statement that condenses an argument into one sentence," and "identifying and distinguishing among critical perspectives on communication." Each of the 6 items were averaged into a global index of perceived confidence in learning outcomes (a = .87).

Procedure

Students were informed that instruments would be used to identify achievement of learning outcomes at the beginning and end of the semester. To ensure consistency in the comparison, students were allotted 20 minutes to complete the instrument for both the pretest and posttest measures respectively. As a component of achieving learning outcome 2, students watched and evaluated a brief video advertisement before completing the instrument. To manage the risk of carryover effects from observing the same video from Timepoint 1-2, the videos varied between timepoints to evaluate how well students could make arguments about novel communication texts. All other materials were replicated between pretest and posttest instruments.

In advance of the assessment, students were informed that responses to the inventory would be deidentified and would not impact their grades. Anonymity was ensured by creating a unique identifier which was subsequently used to match responses in the pretest and posttest assessment instrument. Participant identifiers which could not be matched between pretest/posttest scores or grades were excluded from the analysis.

Coding and Reliability

Learning scores were coded by two investigators and a detailed coding procedure was applied to calculate achievement of points corresponding to learning outcomes across Timepoint 1-2. The coding procedure was developed by the course instructor with strict criteria to evaluate whether learning outcomes were met. Each item was coded dichotomously according to whether the learning criteria were met (1 = met; 0 = unmet) and the scores for the 20 items were aggregated for both the pretest and posttest evaluation. Raters produced independent evaluations for 20% of the sample which was used to determine the degree of coder agreement. Learning scores were computed by subtracting aggregate posttest scores from pretest scores. Cohen's kappa was computed to determine the extent of agreement between coders. Kappas ranged from 0.78 to 1.0, which indicates substantial to perfect agreement.

Results

To evaluate the association of student learning attainment at the end of the semester (Hypothesis 1) and learning across the semester (Hypothesis 2) with grades, Pearson's r was computed. Grades were calculated in aggregate for the final and midpoint of the semester as well as at the granular level for each major exam and term papers. Each grade index was then correlated with post-test scores and learning scores (posttest minus pretest values). Each aggregate and granular grade was significantly positively associated with post-test scores. This yielded support for Hypothesis 1. Similarly, significant positive correlations were observed between learning scores and each grade value except for midpoint grades which were positively associated but marginally non-significant (p = .062). This generally provides support for Hypothesis 2 (see Table 1. for correlations).

Table 1. Correlations between Learning scores and Grades.

Measure	1	2	3	4	5	6	7
1. Post-test scores	_						
2. Learning scores	.715**	_					
3. Final grades	.480**	.304*	_				
4. Midpoint grades	.342**	**.214**	.882**	_			
5. Exam 1	.351**	.246*	.731**	.743**	_		
6. Exam 2	.493**	.314*	.650*	.882.**	.550**	_	
7. Papers	.523**	.331*	.938**	.799**	.768*	.676**	_

Note. * indicates $p < .05$; ** indicates $p < .01$.

An ancillary purpose of this project was to explore possible differences in learning between in-person and online course formats (RQ1). An independent sample t-test was conducted to evaluate whether learning and grades varied across instructional formats. Results indicated that there was a significant difference in post-test scores ($t[66]=3.18$, $p < .005$) with the online section demonstrating significantly higher post-test scores ($M = 16.13$, $SD = 2.24$) compared to the in-person section ($M = 12.52$, $SD = 4.33$). Similarly, the online section exhibited higher learning ($M = 9.50$, $SD = 4.76$), compared to the in-person section ($M = 5.80$, $SD = 4.74$), $t(66) = 2.55$, $p < .05$. Exploring potential differences in online/in-person for grades, an independent sample t-test was conducted for major assignments individually and for grades overall. It was found that there was no significant difference between online/in-person sections across final grades, exams, or papers (all t values < 1.12). While the online section scored significantly higher on both post-test and learning scores compared to in-person, sections did not differ in performance on student grades.

An additional research question sought to explore if students' subjective opinions about course satisfaction and confidence was associated with their observed class performance (RQ2). A Pearson's r test was conducted to identify possible correlations between course satisfaction, confidence, aggregate grades, and individual assignments. It was revealed that there was no significant correlation between satisfaction scores and grades. However, results yielded positive correlations between student self-reported confidence in the course learning outcomes and grades (see Table 2 for correlations). The exception was Exam 2 which was not found to be associated with

confidence (p = .835). Taken together, this suggests that grades were not associated with satisfaction but were positively related to confidence.

Table 2. Correlations between Satisfaction, Confidence, and Grades.

Measure	1	2	3	4	5	6	7
1. Satisfaction	_						
2. Confidence	.688**	_					
3. Final grades	.158	.362**	_				
4. Midpoint grades	.227	.387**	.882**	_			
5. Exam 1	.156	.232*	.731**	.743**	_		
6. Exam 2	- .067	.024	.650*	.882.**	.550**	_	
7. Papers	.097	.302*	.938**	.799**	.768*	.676**	_

Note. * indicates $p < .05$; ** indicates $p < .01$.

Discussion

This study explored several insights to increase understanding of the relative correspondence of grades and student learning. Building on existing studies which targeted a reliable assessment instrument for writing-intensive communication courses (see Boyd, 2010; Boyd & Morgan, 2018), this study explored the extent to which pretest/posttest assessment results were associated with grading systems. The assessment instrument targeted basic writing skills, critical thinking skills, description of basic rhetorical concepts, and recognition of different qualitative methods, some of which may be difficult to assess through grades alone. Importantly, this instrument circumvented the grade motivation associated with traditional grades by grounding the items in course outcomes, reducing the likelihood of score inflation, and avoiding performance incentives. Consequently, the instrument served as a litmus test to evaluate the extent to which grades reflect an independent evaluation of learning.

Amid the many criticisms levied against grading systems in higher education, the findings suggest that grades generally exhibited moderate-large positive association with posttest scores from the end of the semester. This provided some evidence that an independent assessment instrument can affirm and support that grade systems reflect student learning. An alternative explanation is that

the post-test scores may have been strongly associated with grades due to the assessment being close in time to when grades were administered. However, grades were shown to be strongly associated with scores on assignments throughout the semester. Moreover, individual grades which were distal in time to the post-test assessment were among the strongest associations. Taken together, the instrument provides a framework for assessment scholars to evaluate the extent to which grades reflect learning with some evidence affirming an association.

Though posttest scores are a useful index of learning achievement, students do not begin a course as a *tabula rasa*. As courses don't exist in a vacuum, it stands to reason that students may enter the course with some existing knowledge (and some may have taken the course again after an earlier failed attempt). As a result, learning gains across the semester warrant evaluation to provide a holistic understanding of student learning. As with scores at the end of the semester, learning was positively associated with both the end of the semester grade and each assignment individually. The exception was midpoint grades, which were positively associated but marginally non-significant. However, the positive association between grades and learning was more modest compared to posttest scores. One explanation for this is that items targeting learning outcomes 1 and 2 were achieved by most students in Timepoint 1(Outcome 1 = 82%; Outcome 2 = 55%), possibly deflating the value of learning gains.

These results offer a nuanced perspective to existing mixed research on grades and student learning. Many scholars have strongly questioned the extent to which grades reflect deep learning or academic rigor. Campbell and Cabrera (2014) found that self-report inventories of integrative learning, reflective learning, and higher order learning were not associated with GPA. These criticisms are echoed by many scholars who question the academic rigor and relationship to learning (Arum & Roksa, 2011). Other researchers found that student performance depends strongly on motivation, thus questioning the validity of low-stakes assessment on reflecting learning. Though a modest sample, this study found that both individual and aggregate grades were strongly associated with student learning. Moreover, this association was found despite the low-stakes nature of the instrument. This suggests that a pretest/posttest assessment procedure grounded in specific learning objectives can accurately reflect grades and may serve as a fruitful tool to assess student performance for writing-intensive communication courses.

A supporting purpose of the study was to explore whether the assessment instrument and grades varied across online vs. in-person formats. Due to the recent wide-spread adoption of online instruction in response to the COVID-19 pandemic, identifying differences in educational outcomes between instructional formats has become increasingly important. This study found that the online section yielded higher posttest scores and learning compared to the in-person section, while no difference was observed between grades. This reflects emerging research which suggests that educational outcomes are not reduced due to online instruction (Mata, 2021; Welch et al., 2022). This offers alternative evidence to the body of research which has examined the many challenges associated with online instruction (see Kebritchi et al,. 2017). Though the limited sample size warrants careful interpretation, these findings provide evidence that online instruction can meet or exceed in-person formats when it comes to student achievement measured by pretest/posttest assessment indices.

The study further explored the extent to which instructor-based evaluations coincided with student self-appraisals of their performance. While student satisfaction was not associated with grades, the study yielded a general, positive correlation with grades and perceived confidence. This mirrors the findings of Nicholson et al. (2013) which noted an association between student beginning-of-semester confidence and end-of-semester grades. By contrast, the current study evaluated confidence at the end of the semester when grades were likely top-of-mind, thus influencing self-appraisals of confidence. Though further exploration is warranted, this provides

some evidence for the correspondence between student perceived and instructor observed academic performance.

Practical Implications

The results of this study can provide guidance for best practices in assessment and course design. First, Hypotheses 1 and 2 were supported, indicating that student achievement of learning outcomes was positively correlated with course grades. The use of pretest and posttest assessment allows for operationalization of course learning outcomes so that grades are not the only measure of whether learning outcomes were achieved. For the purposes of SOTL, as well as assessment for internal or accreditation purposes, pretest and posttest comparisons provide objective measures of actual student learning—measures that should ideally also be aligned with student grades.

To improve the correlation between student learning and grades, instructors should rely on Backward Design in which desired results drive measures of learning and, consequently, the content of instruction (Wiggins et al., 2005). Accordingly, one feature of this course is the instructor's careful attention to assignment development to ensure that all homework, papers, and exam questions align with at least one SLO. The instructor consistently reviews and revises every aspect of the class, and removes all extraneous content, so that each assignment directly measures students' achievement of SLOs. Only assignments which affirm learning for a specific outcome remain in the class which can lead to grades corresponding to learning. This assessment of SLOs on a granular level affords instructors to identify the specific areas of student need and modify instruction accordingly.

At a time when many scholars are questioning the rigor of college courses, it is more important than ever for instructors to demonstrate whether students are meeting the intended learning objectives. This study indicates that pre/posttests can support the claim that course grades reflect student learning. Assessment instruments should be grounded in specific learning objectives to accurately measure the correlation between grades and achievement of SLOs. Instructors are encouraged to use anonymous, low-stakes assignments for learning assessment to decrease students' motivation to cheat on these tests.

Historically, instructor effectiveness has been measured using end-of-semester Student Evaluations of Teaching, giving students the opportunity to reflect on their experience in the course. Critics have questioned the validity of SETs, arguing that students' responses are impacted by many factors besides the competence of the instructor. Academics may also be motivated to design less rigorous courses with more lenient practices to promote positive feedback on SETs. This study found that, while student perception of learning was positively correlated with course grades, their overall satisfaction with the course was not. This result supports the continued use of SETs for measuring student achievement of SLOs but indicates that student satisfaction with a course may not be an accurate measure of their instructor's teaching abilities.

Expedited by the COVID-19 pandemic, universities have recently been compelled to provide students with more online course offerings, but some scholars have expressed concern that this format decreases course rigor and student achievement. The current study challenges these assumptions with modest evidence suggesting that students may learn *more* in online classes. Although there was no difference in student grades between course modalities, online students exhibited more learning and higher posttest scores than their face-to-face counterparts. These results support the notion that continuation of online course offerings at the college level can produce comparable evidence of learning to traditional course formats.

Limitations and Future Directions

Despite this study's contributions, there are some limitations that must be addressed. First, this study explored the role that course format has on students' grades and learning outcomes. However, because not every in-person student attended each class session, it is possible that they did not fully benefit from all instruction; due to the fluctuation of students under quarantine during the pandemic, the lecture videos were available online for all students and many of the students registered for the in-person class elected to watch online rather than attending the face-to-face class meeting. Future research should explore how the number of absences impacts students' grades and learning, possibly comparing situations where access to lecture and course content is available online and situations where it is not. Additionally, the dataset in this study came from a single course at one public, research-focused university in the United States. It is possible that results would differ if the same data were collected at a different type of institution in another location. Furthermore, because writing is contextual, the same study conducted in a different writing-intensive class may yield different findings. This study should be replicated in other settings to determine how student demographics, institution type and location, and course themes and learning objectives impact results. Replication of this research will also provide guidance on evidence-based best practices for assessing learning in writing-intensive courses, developing assignments that align with learning objectives, and evaluating student work according to intended learning outcomes.

Appendix A

Course Learning Outcomes

1. Write a clear thesis statement that condenses an argument into one sentence. (0 - 3 points)
2. Write and support an argument about a communication text. (0 - 3 points)
3. Explain how rhetoric is foundational for communication. (0 - 4 points)
4. Distinguish among qualitative methods of communication research. (0 - 3 points)
5. Identify and distinguish among critical perspectives on communication. (0 - 3 points)
6. Apply multiple methods to a research scenario (0 - 4 points)

Appendix B

Learning Outcome Items

1. You are completing an online (face-to-face) course, and you've probably also taken classes delivered face-to-face (online) . What do you think—are online classes more effective methods of instruction, or are conventional face-to-face classes? You don't need to make a thorough argument for your position, but in the space below write a complete thesis statement around which you would build an argument answering the question above. (LO1; 3 points possible)
2. In one well-developed paragraph, write an argument below about this short video. (LO2; 3 points possible)
3. Name the person recognized as the most influential thinker in the history of rhetoric and list two of that person's ideas that remain important and used today. (LO3; 3 points possible)
4. Offer one example of the rhetorical concept of identification (it can be a made-up example or an actual one, but it should show clear application of the concept). (LO3; 1 point possible)

5. In one sentence or phrase, define ethnography. (LO4; 1 point possible)
6. What qualitative method studies the basic building blocks of interaction to understand how communication acts (e.g., questioning, telling a joke, complaining) work? (LO4; 1 point possible)
7. What qualitative method believes that the interaction of several people about a question or issue produces richer and more complete information than asking people that same question individually, without the interaction with others? (LO4; 1 point possible)
8. What are two significant differences between a modernist and postmodern perspective? (LO5; 2 points possible)
9. In ideological criticism, what is assumed to be the root or default explanation for how communication texts work? Answer this question in a phrase or a sentence. (LO5; 1 point possible)
10. Imagine that you wanted to study a student singing group at this university. In the space below, offer two different qualitative methods or critical perspectives that you might use to understand this singing group, along with a sentence briefly describing how you could apply that method or perspective. (LO6; 4 points possible)

References

Anderson, D., Imdieke, S., & Standerford, N. S. (2011). Feedback please: Studying self in the online classroom. *International Journal of Instruction, 4*, 3–15.

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

Bazerman, C., Little, J., Bethel, L., Chavkin, T., Fouquette, D., & Garufis, J. (2005). *Reference guide to writing across the curriculum*. Parlor Press.

Boyd, J. (2010). The best of both worlds: The large lecture, writing-intensive course. *Communication Teacher*, 24(4), 229-237. https://doi.org/10.1080/17404622.2010.513992

Boyd, J., & Morgan, M. (2018). A tool instead of a chore: Measuring student learning gains in order to improve instruction. *Communication Teacher, 32*(3), 154-166. https://doi.org/10.1080/17404622.2017.1372596

Campbell, C. M., & Cabrera, A. F. (2014). Making the mark: Are grades and deep learning related?. *Research in Higher Education*, *55*(5), 494-507. https://doi.org/10.1007/s11162-013-9323-6

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

Chametzky, B. (2014). Andragogy and engagement in online learning: *Tenets and solutions. Creative Education, 5*, 813-821. http://dx.doi.org/10.4236/ ce.2014.510095

Ching, Y. H., & Hsu, Y. C. (2015). Online graduate students' preferences of discussion modality: Does gender matter?. *Journal of Online Learning and Teaching*, *11*(1), 31-41.

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*, *31*(1), 16-30. https://doi.org/10.1177/0273475308324086

Cole, M. S., Feild, H. S., & Giles, W. F. (2003). Using recruiter assessments of applicants' resume content to predict applicant mental ability and big five personality dimensions. *International Journal of Selection and Assessment*, *11*(1), 78-88. https://doi.org/10.1111/1468-2389.00228

Cureton, L. W. (1971). The history of grading practices. *Measurement in Education, 2*(4), 1-8.

Dahlgren, L. O., Fejes, A., Abrandt-Dahlgren, M., & Trowald, N. (2009). Grading systems, features of assessment and students' approaches to learning. *Teaching in Higher Education*, *14*(2), 185-194. https://doi.org/10.1080/13562510902757260

de Kleijn, R. A., Mainhard, M. T., Meijer, P. C., Pilot, A., & Brekelmans, M. (2012). Master's thesis supervision: Relations between perceptions of the supervisor–student relationship, final grade, perceived supervisor contribution to learning and student satisfaction. *Studies in Higher Education*, *37*(8), 925-939. https://doi.org/10.1080/03075079.2011.556717

Delucchi, M. (2014). Measuring student learning in social statistics: A pretest-posttest study of knowledge gain. *Teaching Sociology, 42*(3), 231-239. https://doi.org/10.1177/0092055X14527909

Farris, C., & Smith, R. (1992). Writing-intensive courses: Tools for curricular change. In S. H. McLeod & M. Soven (Eds.), *Writing across the curriculum: A guide to developing programs* (pp. 71-86). Sage.

Good, J. M., Osborne, K., & Birchfield, K. (2012). Placing data in the hands of discipline-specific decision makers: Campus-wide writing program assessment. *Assessing Writing*, *17*(3), 140-149. https://doi.org/10.1016/j.asw.2012.02.003

Gunawardena, C. N., & Zittle, F. J. (1997). Social presence as a predictor of satisfaction within a computer-mediated conferencing environment. *American Journal of Distance Education*, *11*(3), 8-26. https://doi.org/10.1080/08923649709526970

Haswell, R. H. (2000). Documenting improvement in college writing: A longitudinal approach. *Written Communication, 17*(3), 307–352. https://doi.org/10.1177/0741088300017003001

Imose, R., & Barber, L. K. (2015). Using undergraduate grade point average as a selection tool: A synthesis of the literature. *The Psychologist-Manager Journal*, *18*(1), 1-11. https://doi.org/10.1037/mgr0000025

Jaschik, S. (2011, January 18). Academically adrift. In *Inside Higher Ed*. Retrieved March 30, 2022, from https://www.insidehighered.com/news/2011/01/18/academically-adrift.

Jewell, R. T., McPherson, M. A., & Tieslau, M. A. (2013). Whose fault is it? Assigning blame for grade inflation in higher education. *Applied Economics*, *45*(9), 1185-1200. https://doi.org/10.1080/00036846.2011.621884

Kebritchi, M., Lipschuetz, A., & Santiague, L. (2017). Issues and challenges for teaching successful online courses in higher education: A literature review. *Journal of Educational Technology Systems*, *46*(1), 4-29. https://doi.org/10.1177/0047239516661713

Kelly, S. (2008). What types of students' effort are rewarded with high marks?. *Sociology of Education*, *81*(1), 32-52. https://doi.org/10.1177/003804070808100102

Kleinman, S. B., Leidman, M. B., & Longcore, A. J. (2018). The changing landscape of grading systems in US higher education. *Perspectives: Policy and Practice in Higher Education*, *22*(1), 26-33. https://doi.org/10.1080/13603108.2017.1279692

Ko, S., & Rossen, S. (2017). *Teaching online: A practical guide.* Routledge.

Limperos, A. M., Buckner, M. M., Kaufmann, R., & Frisby, B. N. (2015). Online teaching and technological affordances: An experimental investigation into the impact of modality and clarity on perceived and actual learning. *Computers & Education*, *83*, 1-9. https://doi.org/10.1016/j.compedu.2014.12.015

Lindblom-Ylänne, S., Parpala, A., & Postareff, L. (2019). What constitutes the surface approach to learning in the light of new empirical evidence?. *Studies in Higher Education*, *44*(12), 2183-2195. https://doi.org/10.1080/03075079.2018.1482267

Lui, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher

education: Motivation matters. *Educational Researcher, 41*, 352–362. https://doi.org/10.3102/0013189X12459679

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

Martin, F., Ritzhaupt, A., Kumar, S., & Budhrani, K. (2019). Award-winning faculty online teaching practices: Course design, assessment and evaluation, and facilitation. *The Internet and Higher Education*, *42*, 34-43. https://doi.org/10.1016/j.iheduc.2019.04.001

Mata, J. R. (2021). How to Teach Online? Recommendations for the assessment of online exams with university students in the USA in times of pandemic. *IJERI: International Journal of Educational Research and Innovation*, (15), 188-202. https://doi.org/10.46661/ijeri.5003

McMorran, C., Ragupathi, K., & Luo, S. (2017). Assessment and learning without grades? Motivations and concerns with implementing gradeless learning in higher education. *Assessment & Evaluation in Higher Education*, *42*(3), 361-377. https://doi.org/10.1080/02602938.2015.1114584

Nicholson, L., Putwain, D., Connors, L., & Hornby-Atkinson, P. (2013). The key to successful achievement as an undergraduate student: Confidence and realistic expectations?. *Studies in Higher Education*, *38*(2), 285-298. https://doi.org/10.1080/03075079.2011.585710

Nusche, D. (2008), "Assessment of Learning Outcomes in Higher Education: a comparative review of selected practices", OECD Education Working Papers, No. 15, OECD publishing, © OECD. https://doi.org/10.1787/244257272573

Oncu, S., & Cakir, H. (2011). Research in online learning environments: Priorities and methodologies. *Computers & Education*, *57*(1), 1098-1108. https://doi.org/10.1016/j.compedu.2010.12.009

Pynes, C. A. (2014). Seven arguments against extra credit. *Teaching Philosophy*, *37*(2), 191-214. https://doi.org/10.5840/teachphil20144414

Qenani, E., MacDougall, N., & Sexton, C. (2014). An empirical study of self-perceived employability: Improving the prospects for student employment success in an uncertain environment. *Active Learning in Higher Education, 15*(3), 199-213. https://doi.org/10.1177/1469787414544875

Rojstaczer, S., & Healy, C. (2012). Where A is ordinary: The evolution of American college and university grading, 1940-2009. *Teachers College Record*, *114*(7), 1-23. https://doi.org/10.1177/016146811211400707

Sadler, D. R. (2013). The futility of attempting to codify academic achievement standards. *Higher Education,* 67(3), 273–288. https://doi.org/10.1007/s10734-013-9649-1

Schinske, J. & K. Tanner. (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

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, *143*(6), 565. https://doi.org/10.1037/bul0000098

Slomp, D. H. (2012). Challenges in assessing the development of writing ability: Theories, constructs and methods. *Assessing Writing*, *17*(2), 81-91. https://doi.org/10.1016/j.asw.2012.02.001

Spada, M. M., & Moneta, G. B. (2012). A metacognitive-motivational model of surface approach to studying. *Educational Psychology*, *32*(1), 45-62. https://doi.org/10.1080/01443410.2011.625610

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598-642. https://doi.org/10.3102/0034654313496870

Stroebe, W. (2016). Why good teaching evaluations may reward bad teaching: On grade inflation and other unintended consequences of student evaluations. *Perspectives on Psychological Science, 11*(6), 800-816. https://doi.org/10.1177/1745691616650284

Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, *42*(4), 276-294. https://doi.org/10.1080/01973533.2020.1756817

Swarat, S., Oliver, P. H., Tran, L., Childers, J. G., Tiwari, B., & Babcock, J. L. (2017). How disciplinary differences shape student learning outcome assessment: A case study. *AERA Open*, *3*(1), 1–12. https://doi.org/10.1177/2332858417690112

Tippin, G. K., Lafreniere, K. D., & Page, S. (2012). Student perception of academic grading: Personality, academic orientation, and effort. *Active Learning in Higher Education*, *13*(1), 51-61. https://doi.org/10.1177/1469787411429187

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, *54*, 22-42. https://doi.org/10.1016/j.stueduc.2016.08.007

Welch, J. R., Roland, C., & Boyd, J. (2022). Condensed and uploaded: Comparing student learning outcomes in a condensed online summer class with outcomes in a full-length semester face-to-face class. *Summer Academe: A Journal of Higher Education, 14*. http://doi.org/10.25894/sa.123

Wiggins, G., Wiggins, G. P., & McTighe, J. (2005). *Understanding by design*. Ascd.

Zheng, Y., & Yu, S. (2019). What has been assessed in writing and how? Empirical evidence from assessing writing (2000–2018). *Assessing Writing*, *42*, 1-11. https://doi.org/10.1016/j.asw.2019.100421