ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5822
https://doi.org/10.24059/olj.5822

Research quality moderates the relationship between student motivation and academic performance in online and blended courses: A meta-analysis

Andrew Walker https://orcid.org/0000-0001-8693-9921
Kaelyn N. Loveless https://orcid.org/0009-0006-8099-3824
Utah State University
Naomi R. Aguiar https://orcid.org/0000-0002-3139-9654
Oregon State University
Yu-Chun Kuo https://orcid.org/0000-0002-7147-9798
Rowan University
Raechel N. Soicher https://orcid.org/0000-0002-2142-625X
Massachusetts Institute of Technology
Jessica Resig https://orcid.org/0000-0003-2706-8446
Applied Research Institute

Abstract

Prior meta-analytic work suggests the relationship between student motivation and academic performance among in-person learners is strong; however, evidence is weaker for college students learning in online, digital, and distance education (ODDE). This could be due to a variety of factors, including the rigor and quality of existing research, which has not been systematically accounted for in (ODDE) motivation reviews since 2008. The primary objective of this expanded meta-analytic review was to examine measurement quality and methodological quality first as moderators and then as adjustments in exploring the relationships between motivational constructs or synchronicity and academic performance in ODDE courses. Informed by DISCAR and adapted from PRISMA, we searched nine databases of studies published between 1990 and 2025 and conducted a secondary search of references, ultimately analyzing 64 studies and 117 outcomes that met inclusion and exclusion criteria. Both measurement quality and methodological quality were moderators of the relationship between student motivation and academic performance in online and blended courses. Adjusting for significant quality differences, overall, modest correlation values ($r = 0.14$) overestimated adjusted relationships ($r_{adj} = 0.10$). Despite challenges with some instances of missing data and low interrater reliability, these results indicate that motivational factors related to student success may be overstated when research quality is taken into account. Variations were found in both motivational constructs as well as synchronicity, suggesting that what is measured and how learners interact make a difference. These results have implications for future ODDE research, pedagogical designers, and policymakers.

Keywords: meta-analysis, online learning, blended learning, motivation, research quality

Corresponding author: Dr. Andrew Walker, andy.walker@usu.edu
This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 The Author(s).

Introduction

In higher education, online and blended course modalities are of increasing importance to institutional success. Online and blended course offerings provide the opportunity to expand access to higher education beyond the "traditional" in-person learner. This expanded access to higher education not only benefits diverse learners everywhere; it benefits academic institutions, many of which are facing emergent enrollment challenges (National Center for Educational Statistics, 2024). Although these and other benefits of online and hybrid modalities are well documented (Lucas & Vicente, 2023), so too are its challenges (Kebritchi et al., 2017; Lucas & Vicente, 2023) and complexities, including variations in synchronous and asynchronous instruction (Johnson et al., 2022). One particular concern is student academic performance, as it is associated with retention and graduation rates (Gardner et al., 2024), which tend to be lower in online programs (Bawa, 2016).

Student motivation is currently of high interest in online and hybrid courses. In addition to a large volume of research, it plays a prominent role in conceptual frameworks like Community of Inquiry (CoI) for online settings (Shea et al. 2022). This focus coincides with investigations of in-person learners, where the relationship between student motivation and academic performance is strong (Honicke et al., 2020; Richardson et al., 2012). These findings with in-person learners have contributed to the widely held belief that motivation is one of the most important predictors of college students' academic performance (Honicke et al., 2020). However, our recent meta-analytic review indicates that the relationship between student motivation and academic performance is more modest for online distance and digital education (ODDE), and accounts for little of the variance in academic performance (Walker, Aguiar, Soicher, Kuo, & Resig, 2024). This appears true across a broad range of motivational constructs, including self-efficacy, attributional style, achievement goal orientation, self-determination and task value.

It is possible that these modest relationships are partially explained by a range of moderating factors. Methodological quality clearly matters, both to the academic community and as context for findings. Early reviews of motivation and learning express concerns regarding study quality (Bernard et al., 2004; Multon et al., 1991; Perkowski, 2013; Storrings, 2005). Reductions in study quality are associated with reductions in effect sizes for in-person learning (Macnamara et al., 2023) and for ODDE learners (Bernard et al., 2004). Measurement quality also matters in reviews of motivation. Some reviews code for and include reliability of measurement in their analyses (Richardson et al., 2012) but not validity. Others, make strong contributions to validity in specific constructs like achievement goals (Huang, 2012) but do not address reliability. Investigations of ODDE literature that address study quality only cover literature to 2006 (Bernard et al., 2009) and measurement quality only covers to 2008 (Huang, 2012). Of the approaches to addressing quality (Fromm et al., 2025), existing reviews have employed selection criteria (Storrings, 2005), and novel coding (Multon et al., 1991). None have used an establish quality rubric (Katrak et al., 2004; Rosella et al., 2016a; Sterne et al., 2019; Whiting et al., 2016) or comprehensively coded for methodological quality alongside measurement reliability and validity. Finally, asynchronous and synchronous interactions clearly matter in this literature and are associated with variations in effect sizes (Bernard et al., 2004) and persist when coding for a mixture of the two (Bernard et al., 2009). Like study and

measurement quality, reviews of the ODDE literature have not coded for synchronicity for almost two decades of primary research. A recent critical review (Krebs-Lazendic et al., 2026) indicated that synchronicity within online courses may be an important factor in the focused area of instructor presence and academic performance, warranting replication newer results in broader ODDE literature.

The goal of the current meta-analytic review was to expand the previous work of Bernard et al. (2009) and Richardson et al. (2012) by comprehensively examining methodological quality and measurement quality as moderators of the relationship between student motivation and academic performance in online and blended college-level courses. Importantly, not only was it crucial to assess the impact of methodological and measurement quality on the relationship between motivation and academic performance in ODDE research, it was important to replicate and extend adjustments for measurement reliability (Richardson et al., 2012) to include adjustments for measurement validity and methodological quality in the examination of moderators for motivational constructs and synchronicity.

Existing Reviews

In an effort to be comprehensive, we examined 19 historical, scoping, systematic or meta-analytic reviews overlapping with post-secondary education. About half drew from in-person research (Howard et al., 2021; Hulleman et al., 2010; Lazowski & Hulleman, 2016; Macnamara et al., 2023; Multon et al., 1991; Richardson et al., 2012; Senko & Dawson, 2017; Theobald, 2021) and the rest summarized the ODDE literature (Bernard et al., 2004, 2009; Huang, 2012; Lee, 2017; Mahmud et al., 2022; Martin et al, 2022; Perkowski, 2013; Storrings, 2005; Walker et al., 2024; L. Wang, 2022; Zhao et al., 2005). Variation between in-person and ODDE learners across reviews is stark. The relationship between variations of self-efficacy and academic performance in Richardson et al. (r = 0.31 to 0.59) are clearly stronger than the aggregate self-efficacy for ODDE contexts, Walker et al. (2024) (r = 0.14). Further exploration of moderators is clearly needed.

Not only were meta-analyses addressing primary research quality sporadic (Bernard et al., 2004, 2009; Macnamara et al., 2023; Multon et al., 1991; Storrings, 2005), but none of them used an establish quality coding rubric. Reviews addressing methodological quality ranged from simple reporting (Multon et al., 1991) to using publication in peer reviewed journals as a form of quality control (Richardson et al., 2012), to analyses indicating that variations in quality do impact effect sizes (Bernard et al., 2004; Macnamara et al., 2023). For example, Macnamara et al. (2023) found that significant effect sizes became non-significant when focusing on acceptable or best available study quality. Existing reviews note challenges of insufficient data reporting (Multon et al., 1991) including methodological quality, making recommendations to stakeholders challenging (Bernard et al., 2004). Other quality concerns range from methodological rigor (Storrings, 2005) to convenience samples and treatment variations (Perkowski, 2013). Despite clear concerns, the most recent review incorporating quality in the analysis of motivation for ODDE learners (Bernard et al., 2009) only covers studies as recent as 2006. All of these reviews

use novel coding for methodological features, bypassing modern quality rubrics (Rosella et al. 2016).

Measurement quality is not routinely addressed within motivation reviews. Some use measurement rigor as a selection criterion (Howard et al., 2021; Storrings, 2005). Others, working in the niche area of Achievement Goals report reliabilities across scales (Hulleman et al., 2010). Rare reviews (Richardson et al., 2012) report and adjust for internal construct reliability of both predictors and reliability of learning outcomes. Even studies that address reliability in some fashion fail to address validity (Hulleman et al., 2010; Richardson et al., 2012). Among ODDE reviews, the only analysis of measurement focused on discriminant and criterion validity of achievement goals as moderators (Huang et al. 2012) and only covered studies through 2008. For discriminant validity, Huang et al. (2012) found four factor models of achievement goals had the highest correlation values with learning. A meta-regression for criterion validity determined that 6% of the variance in academic achievement could be accounted for when using all four sub-constructs. Clearly, there is a need to examine literature at the intersection of motivation and ODDE using an established methodological quality rubric and comprehensively addressing measurement reliability and validity for a wide range of motivational constructs, covering two decades of recent literature.

Online and Blended Synchronicity

Much like selecting a quality coding rubric for educational research, definitions of synchronous, asynchronous and mixed interactions can be challenging. Some definitions center on interaction in combination with group size. At one extreme, online classes are synchronous when they scale the in-person classroom to multiple sites. Asynchronous interactions are either individual or small group based and utilizing technologies like email that allow for separation of time (Bernard et al., 2004). Refined definitions acknowledge directionality as a difference—reading an article is a one-way transfer and fundamentally asynchronous, contrasted with synchronous exchanges like the bi-directional transfer of a chat or Zoom meeting (Bernard et al., 2009). Our definition borrows from the consensus-built work of Johnson et al. (2022), where synchronous learning requires fixed, real-time exchanges, whether in-person or virtually, and asynchronous learning provides students with flexibility to access learning experiences when and where it works for them.

Based on both past and recent reviews (e.g. Bernard et al. 2004; Krebs-Lazendic et al. 2026), course synchronicity may be as important as the nature of the motivational construct. Bernard et al. (2004) examined differences between distance education (DE) and classroom instruction (CI) across affective, performance, and completion outcomes. Splitting results out by asynchronous and synchronous contexts, authors found variations in effect size for academic performance, as well as affective and course completion outcomes. Expanding their prior coding to include mixed synchronicity, Bernard et al. (2009) explored variations in online learning coding for type of interaction (e.g., student-student) and mean difference based on strength of interaction (minimal, moderate, high). Although they did not find significant differences in effect sizes between synchronous and asynchronous courses, they did find that the strength of

interactions in asynchronous courses predicted academic performance; this was not the case for mixed or synchronous courses, suggesting that the strength of interactions is of particular importance when students engage asynchronously with course content. Note that both of these meta-analyses explore mean differences—either between online and in-person settings (Bernard et al., 2004) or between interaction strength variations in online classes (Bernard et al., 2009). Given that other research examining the effects of synchronicity on academic performance remains limited (Krebs-Lazendic et al., 2026), extending early mean difference reviews to explore the moderating effect of synchronicity in motivation relationships is needed.

Motivational Constructs

Building from social cognitive theory to the constructs and subconstructs of student motivation is another review challenge. Motivation is a multidimensional latent variable that consists of several constructs including but not limited to self-efficacy, achievement goal orientation, self-determination, task value and attributional style (see Table 1 for well-established definitions).

Table 1

Select motivational constructs and their definitions

Construct	Definition
Self-efficacy	A person's belief in their ability to organize and execute the tasks required to achieve a desired level of performance, derived from Bandura's Social Cognitive Theory (1986, 2006) and thought to be domain specific.
Achievement Goal Orientation	"A future focused cognitive representation" (Hulleman et al. 2010, p. 423) that guides behavior either towards or away from a specific outcome; measured across dimensions of mastery vs. performance goals and approach vs. avoidance goals (Bardach et al., 2020; Elliot & McGregor, 2001).
Self-Determination	Intrinsic motivation involves pursuing a goal because of its perceived internal value (e.g., for interest or fun); extrinsic motivation involves pursuing a goal that has outcomes separate from the activity (e.g., an external reward, Ryan & Deci, 2000, 2020)
Task Value	Operationalized as attainment value (importance of doing well on a task), intrinsic value of the task, utility value (how a task relates to current and future goals), and cost (Eccles & Wigfield, 2020).
Attributional Style	A perception of successes and failures, made up of three components: (1) locus of causality (internal or external causes), (2) stability (constant or variable), and (3) controllability (Weiner, 2010).

Theoretical Framework for Motivation

There are several intersecting theories in ODDE and motivation literature, such as the Community of Inquiry Framework (Garrison, 1999, 2007), which incorporates self-regulated learning and self-efficacy in learner presence (Chang et al., 2026). The primary interest for this

review is motivation. We adopted the review stance of Lazowski and Hulleman (2016) that motivation is complex, multi-dimensional, and in the context of educational research shares common roots in social cognitive theory (Bandura, 1986, 2006). Some motivational constructs, such as self-efficacy are featured in social cognitive theory. Others, such as achievement goal theory lean heavily on student perceptions and thus a social-cognitive orientation to illuminate goal structures (Ames & Archer, 1988).

Motivation, Measurement, and Synchronicity

In the extant literature specific to ODDE, much of the research has focused on self-efficacy as it relates to academic performance. For example, in our recent meta-analysis, nearly half of the outcomes included in the analyses examined the relationship between self-efficacy and academic performance (Walker et al., 2024). No other single motivational construct (e.g., task value, attributional style) was represented as frequently in empirical research. That could be a function of reviews with in-person learners (Richardson et al., 2012) where the strongest relationships are between self-efficacy scales and academic performance. Note that at least with Richardson et al. (2012), subconstructs of goal orientation accounted for more than half (n = 164) of the total outcomes.

It is possible that the focus on self-efficacy partially stems from the robust empirical testing of instruments that are available to measure it. The Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, 1991) is one such measure that has been rigorously vetted in the literature (de Araujo et al., 2023; Hilpert et al., 2013; Pintrich et al., 1993) and is frequently used. Although this instrument was designed to measure a range of motivational constructs, the Self-Efficacy for Learning and Performance subscale of the MSLQ is often used to measure motivation among online and blended learners (e.g., Alkış & Temizel, 2018; Goulão, 2014; Joo et al. 2013). However, the value of this tool in assessing motivation in online courses has been called into question in past research. Cho and Summers (2012) found that the original factor structure of the MSLQ was not supported in students taking asynchronous online courses. This could partially account for the modest overall relationship we found between self-efficacy and academic performance in our recent meta-analytic review (Walker et al., 2024). However, we also found some variation in the overall effect size, with stronger correlations among self-efficacy for learning in general, and weaker relationships with one of several technology contexts (e.g., online, computer, or LMS self-efficacy). This variation could also be due to the range of approaches in measuring self-efficacy in the literature. While some authors used the original self-efficacy subscale in the MSLQ to measure motivation (e.g., Alkış & Temizel, 2018; McPhaul-Moore, 2013; Savoji, 2013), others modified the original subscale (Hodges, 2008; Liu et al., 2019), and some authors developed their own measurement instruments (Kim, 2012; Martin et al., 2010; Warren et al., 2020). The example of self-efficacy alone highlights the potential for measurement decisions, including the synchronicity of classroom experiences to impact the correlation between motivation and academic performance in online and blended settings. Clearly, an updated review that not only explores the measurement space, but contextualizes findings related to motivational constructs and synchronicity after accounting for measurement quality and methodological quality is needed.

Assessing Study Quality

In meta-analytic reviews, measurement reporting quality is one of several variables best paired with an overall assessment of study quality. Appraisal of study quality, which often looks for reliability and validity of study methods, sampling, results reporting, replicability, and generalizability, is critical for evaluating the relevance and trustworthiness of a study (Rosella et al. 2016). Beyond ensuring a study and its results are trustworthy and relevant, prior meta-analytic work has shown that methodology quality accounts for considerable amounts of variation in effect size (Bernard et al., 2004). Unfortunately, extant reviews cite issues such as high attrition, large amounts of missing data, and differences in course content, duration, and delivery between treatment and control groups as contributors to low quality studies (e.g., Bernard et al. 2004; Huang, 2012; Perkowski, 2013). In a three-round meta-analysis study evaluating methodological quality control in intervention designs (no control, minimum standards, and best available standards), Macnamara et al. (2023) found that failure to account for study quality leads to inflated effect sizes. Additionally, of the 63 studies and 96 effect sizes included in their first-round analysis, only 6 studies and 10 effect sizes met more than 60% of the 10 relevant best practices criteria for third-round analysis. The systematic association between study quality and effect sizes, coupled with the scarcity of methodological quality in the literature, reinforces the importance of quality-controls in analyses meant to advise practitioners and policymakers on best practices.

Though study quality has been shown to impact effect sizes, few meta-analyses of academic achievement in ODDE have accounted for study quality in their analysis or discussion of limitations. Further, when quality is evaluated, authors utilize varying levels and types of criteria to support their assessment. For example, Bernard et al. (2004, 2009) coded for the type of publication, selection bias, treatment-control group equivalence, and attrition while Macnamara et al. (2023) coded for study group similarities, random assignment, a priori power analysis, clear design, transparent and complete results, and reporting of conflicts of interest. Another shortcoming of extant reviews is the exclusion of observational research from analysis. To the best of our knowledge, extant reviews that account for study quality focus primarily on experimental or interventional studies. The shortage of common frameworks or tools to conduct quality appraisals adapted to the wide range of educational research (Fromm et al., 2025) may be a contributor to the shortage of reviews that acknowledge study quality.

While less prominent in educational research, quality appraisal tools have a foothold in evidence-based medicine (EBM), where many tools were developed with a focus on risk of bias (Katrak et al., 2004). Such tools include AMSTAR, a measurement tool to assess systematic reviews (Shea et al. 2007); RoB 2, a revised Cochrane risk-of-bias tool for randomized trials (Sterne et al., 2019); and ROBIS, a tool to assess risk of bias in systematic reviews (Whiting et al., 2016), among others. Quality appraisal tools share at least one limiting factor in their potential usefulness: the inability to adapt to different settings, contexts, and study designs. Rosella et al. (2016) proposed the development of a meta-tool that extends to a wide range of applicability and flexibility, such as to observational studies, as well as maintain a rigorous and transparent process to carry out a detailed assessment. This tool, the Meta Quality Appraisal Tool

(MetaQAT), was developed and validated for the evaluation of a wide range of research applications and practices.

The MetaQAT framework encompasses four domains of quality, including relevancy, reliability, validity, and applicability. The assessment of relevancy focuses on evaluating the degree to which a topic is relevant to the issue that is being investigated. The assessment of reliability focuses on a thorough examination of the quality and transparency of the research, including a clear presentation of a study, the description of research methodology, results, and the inclusion of ethics procedures, etc. In terms of the assessment of validity, it examines both the internal validity and external validity to address the likelihood and magnitude of error or bias in the study; for example, the appropriateness of the study methodology for the scope of research, explicit conclusions, etc. The assessment of applicability evaluates whether the generated results can be applied within the scope of the field or discipline being studied. In the absence of a flexible quality appraisal tool in educational research, the current review adopted a modified MetaQAT to narrow the literature gap examining the influence of study quality on the magnitude of study effects and results in experimental, quasi-experimental, and observational research. In our study, we coded for methodological reliability, validity, and applicability. Relevancy, as outlined in MetaQAT, was addressed through our inclusion and exclusion criteria.

Research Questions

Based on existing reviews, it is clear the quality of primary ODDE research matters. Primary research suggests in-person and ODDE settings are not the same and can impact measurement models (Cho & Summers, 2012). Early review efforts to account for measurement reliability (Richardson et al. 2012), make contributions to measurement validity (Huang, 2012), and explore study quality (Bernard et al., 2004), but have not been replicated in modern ODDE reviews and have never been tackled comprehensively regardless of setting. It is also clear there are differences in the relationship between motivation and academic performance between in-person (Richardson et al., 2012) and ODDE learners (Walker et al., 2024). Regardless of setting, motivational construct is an important moderator in these relationships. In addition, there is a clear need to explore synchronicity of ODDE class interactions based on prior mean difference reviews (Bernard et al, 2004 & 2009) and a recent critical review of the literature (Krebs-Lazendic et al. 2026). Combining these approaches, a replication of Richardson et al. (2012) is needed that includes a comprehensive examination of and adjustments for study quality, an exploration of a broad range of motivational constructs, and an examination of synchronicity levels in online courses.

The current meta-analytic review addresses the following objectives and research questions in ODDE college-level courses:

Objective$_1$: Explore the impact of measurement and methodological quality on studies of motivation and learning performance.

RQ$_1$: To what extent does measurement reporting quality (validity and reliability) moderate the relationships between motivation and learning performance?

RQ_2: To what extent does methodological quality (via MetaQAT and attrition) of primary research moderate the relationships between motivation and learning performance?

$Objective_2$: Explore synchronicity and motivational constructs as moderators, using research quality adjusted effect sizes.

RQ_3: Adjusting for measurement reporting and methodological quality, to what extent does class synchronicity (*asynchronous*, *synchronous*, *mixed*) moderate the relationships between motivation and academic performance?

RQ_4: Adjusting for measurement reporting and methodological quality, to what extent does motivational construct (*self-efficacy, attributional style, achievement goal orientation, and self-determination and task value*) moderate the relationships between motivation and academic performance?

Methods

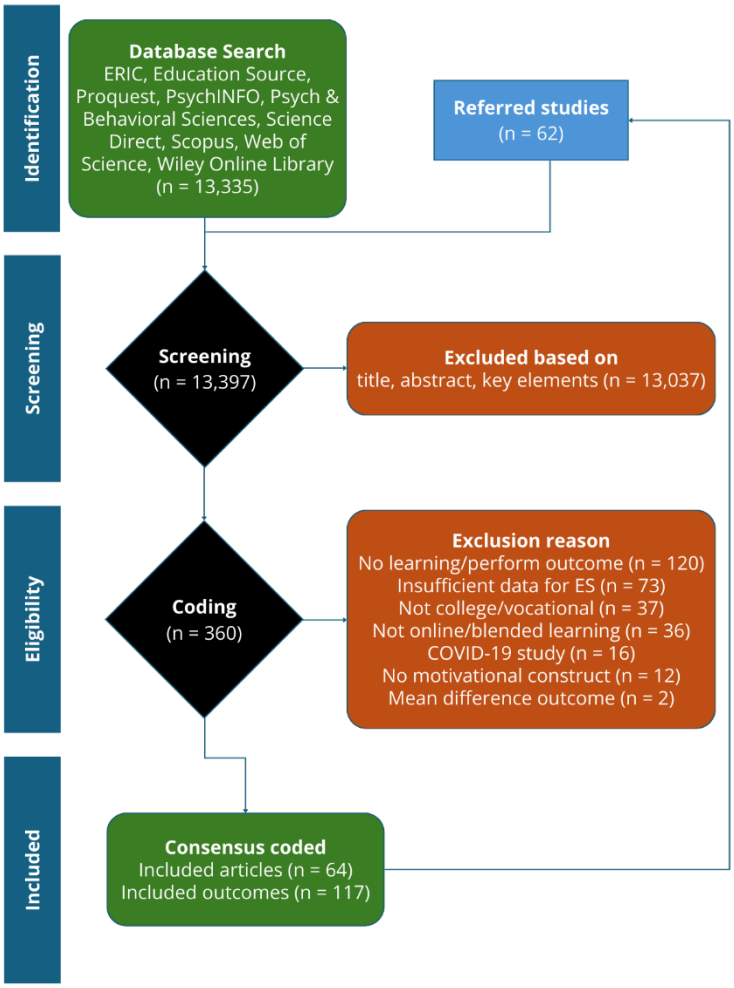
To address our research questions, we employed a systematic review and meta-analysis of ODDE literature examining student motivation and academic performance. Our study protocol was informed by the DISCAR (designing, including/excluding, screening, coding, analyzing, reporting) process (West & Martin, 2024), which identifies best practices for rigorous reviews. Our methods focus on all sections except for reporting, which is addressed in the results section.

Design

Our study and question design was driven by a desire to replicate and extend an exploration of motivational constructs (Richardson et al., 2012) and synchronicity (Bernard et al., 2004, 2009) after exploring and accounting for a broad range of motivational constructs (Walker et al., 2024) and a broad range of research quality in ODDE settings. We began with a scoping review to determine meaningful questions and feasibility of the study, assessing 19 reviews of post-secondary education. Through this scoping review, we identified three major gaps in post-secondary and ODDE literature: 1) inconsistencies in measurement reporting; 2) variation in methodological rigor assessment; and 3) an updated examination of synchronicity. In accordance with existing guidelines, an adapted PRISMA (Page et al., 2021) flow chart (Figure 1) was used for search transparency.

Figure 1

Literature search, referrals, and screening



Inclusion and Exclusion criteria

An early and crucial intersection of quality is selection of the range of methodologies included in the review (Lipsey & Wilson, 2001). Because of our interest in determining the impact of study quality, we cast a wide net for inclusion criteria: 1) Peer reviewed and gray literature; 2) Any study design (descriptive, pre-experimental, quasi-experimental, experimental) as long as it was empirical and quantitative; 3) Measured-attributional style, self-determination, self-efficacy, achievement goals, or task value; 4) Reported a relationship (*r*) between motivation and academic performance; 5) Published from 1990-2025; and 6) Used the English language. Studies were excluded if they: 1) Did not report a learning or performance outcome (most often due to it being a self-report such as anticipated grade); 2) Lacked enough information to calculate a correlation effect size; 3) Had either a mix of learners or education levels that did not align with the college/vocational sampling frame; 4) Were not online or blended learning

experiences; 5) Collected data during COVID-19, defined as any time during 2020; 6) Discussed but failed to report a motivational construct; or 7) Reported results rooted in mean differences instead of relationships.

Replicating the search from our most recent review (Walker et al., 2024), nine academic databases (Education Source, ERIC, PsychInfo, Psychological and Behavioral Sciences, ProQuest, Science Direct, Scopus, Web of Science, and Wiley Online Library) were searched for literature at the intersection of motivational constructs and variations of online teaching modes. Specifically, our search terms were:

(internality OR internal locus of control OR external locus of control OR globality OR stability OR attributional style OR academic locus of control OR optimism ORself-efficacy OR intrinsic motivation OR extrinsic motivation OR goal orientation OR achievement goal) AND (online learning OR distance learning OR web-based learning OR blended learning).

When available, limiters were added for English language, research manuscripts, post-secondary, applied related words, for all available fields encompassing 1990 through 2025. Our search process was inclusive of both peer reviewed journal articles and gray literature including conference papers, research reports and dissertations. Database results were supplemented by referrals from included studies, as well as systematic and narrative literature reviews.

Screening

After the data were extracted from databases, the citations were shared using Zotero and all search results were aggregated in Rayyan.ai (Ouzzani et al., 2016). Rayyan.ai was used to deduplicate results and track pre-pass screening assignments between a team of six screeners. Each study went through a screening of the title, abstract, and in cases of ambiguity about measuring motivation or reporting academic performance, a brief check of the full text. Screeners evaluated exclusion decisions based on key words related to the exclusion criteria, such as COVID; primary, secondary or graduate education; qualitative studies; self-reported performance; and MOOC studies. Generative AI tools were not used in the search, screening, or coding process, nor in the analyses and reporting.

Coding

Following the screening phase, a three-round coding sequence of first-pass coding, second-pass coding, and consensus coding was followed by a team of six researchers. Coding decisions were made according to a live code book which was developed and maintained by the research team. The code book began with deductive, a priori codes based on existing theory or frameworks. Occasionally, inductive coding was necessary when borderline or novel cases emerged in the literature. These borderline or novel cases were documented in the code book with consensus decisions to help ensure future decisions were made in a consistent manner.

Coding included (1) article citation information; (2) class interactions (*synchronous*, *asynchronous*, *mixed*, or *unclear*); (3) motivational construct; (4) sample size; (5) type of correlation; (6) correlation; (7) study attrition based on percentage of sampling frame; (8) quality

of measurement validity reports (*strong*, *attempt*, *none*); (9) quality of measurement reliability reports (*strong*, *attempt*, *none*); and (10) study quality (*good* or *bad/unclear*) utilizing the MetaQAT tool (Rosella et al. 2016).

Promising studies were first-pass coded and rejected if they failed to meet inclusion criteria. Studies that were accepted after a first-pass code were second-pass coded by another researcher. Independent coding was followed by a consensus round where disagreements were resolved including final decisions about inclusion, exclusion, and coding. The resolution of disagreements among coders was centered around finding different information to justify coding choices rather than differences in applying the coding guide. Data from the consensus coding was used in the final analysis.

Analysis

Analysis and reporting were supported through STATA 18.0. Base STATA functions for meta-analysis were supplemented by community contributed functions in MAIS and ROBMETA.

Preliminary Analyses

The following analyses are provided to promote transparency or because they were supportive of analyses that directly addressed research questions. Data extraction for all included studies is available in Appendix A.

Inter-coder reliability

Reliability scores (see Table 2) were not based on a sample, but a comprehensive analysis of all independent first- and second-pass coding. We used Krippendorff's Alpha (2004) for interval, nominal, and ordinal data and Cohen's Kappa (Cicchetti & Feinstein, 1990) for boolean variables (*good* or *bad/unclear*).

Table 2

Inter-rater reliability of all first- and second-pass coding

Variable	reliability	method	data type
Effect size (r)	0.96	alpha	interval
Measurement Reporting Quality			
Validity (*strong*, *attempt*, *none*)	0.90	alpha	ordinal
Reliability (*strong*, *attempt*, none)	0.84	alpha	ordinal
Study Quality (good, bad/unclear unless otherwise noted)			
Reliability: Clear presentation	0.70	kappa	boolean
Reliability: Clear methods and results	0.57	kappa	boolean
Reliability: Ethics procedures described	0.76	kappa	boolean

Variable	reliability	method	data type
Validity: Methods align with scope of research	0.67	kappa	boolean
Validity: Free from bias	0.61	kappa	boolean
Validity: Explicit and transparent conclusions	0.67	kappa	boolean
Applicable to setting (higher education)	0.47	kappa	boolean
Confident in findings	0.79	kappa	boolean
Attrition for sample (0 - 1.00; *missing*)	0.86	alpha	interval
Moderators			
Synchronicity (*asynch*, *synch*, *mixed*, *unclear*)	0.75	alpha	nominal
Motivational construct (thirteen options)	1.00	alpha	Nominal

Most of the Krippendorff's alpha values were at or near acceptable reliability of > .80 (Marzi et al., 2024). For Cohen's Kappa, inter-rater reliability was weak at < 0.59 to moderate at < 0.79 (McHugh, 2012).

Outlier detection

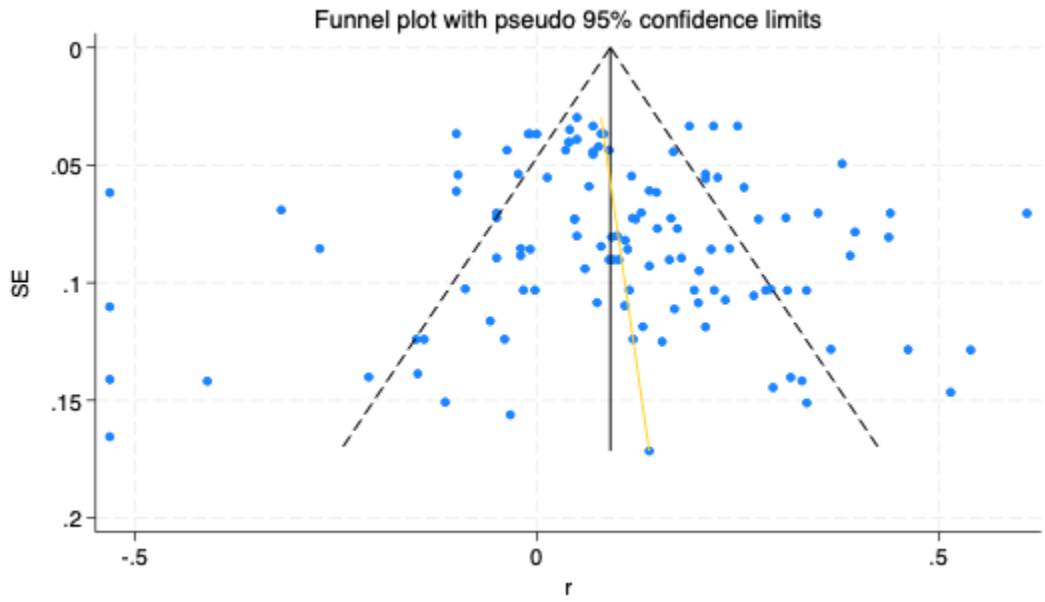
Three negative outliers were detected (Liu et al., 2019; Stephen et al., 2020; Tladi, 2017), prior to any adjustments. Rather than drop meaningful data, the values for each were trimmed to -.411, which is the lowest effect size that is not an outlier. Note these same three outliers were found before and after corrections for quality.

Publication bias

Publication bias using Egger's test was not significant $t(116) = 1.38$, $p = 0.17$ before trimming outliers. After trimming outliers as well as after adjustments for study quality, there was still no support for publication bias $t(116) = 0.79$, $p = 0.43$. A funnel plot of trimmed outlier and quality adjusted effect sizes by standard error (see Figure 2) also suggests there was no publication bias.

Figure 2

Visualization of standard error and adjusted effect size estimates



Multiple outcomes

With a total of 117 outcomes across 64 studies, there were clearly nested data. On average, there were 1.8 outcomes per study, with a range of a single outcome to 8 outcomes. Keeping these data was important to observe the full range of quality and also to keep information about moderators. For example, Bahçekapili and Karaman, (2020) examined the relationships between self-efficacy ($r = 0.14$), external ($r = 0.16$) and internal ($r = 0.03$) locus of control to academic performance. These not only represent different motivational constructs; they represent effect sizes that are both above and indistinguishable from 0. Rather than drop outcomes, we used Robust Variance Estimation (Hedges et al., 2010) to determine the degree to which the study of origin impacts effect size estimates of individual outcomes. Using adjusted effect sizes, the difference between assuming that no variation (*rho* = 0.00) was from the study of origin and all variation (*rho* = 0.99) was from the study of origin was minimal. Specifically, there was a 0.01 change in the overall effect size for the differences between no variation and all variation attributed to the study of origin; there was no change for quality adjusted overall effect size. As a result, outcomes were treated as if they were independent.

Academic Performance

Given the self-reported nature of most measures of motivation, our interest in academic performance was in scored performances including tests, quizzes, or class grades. We also included class completion and persistence (e.g., returning for additional coursework; Stephen, 2020). We excluded self-report of performances like grades (e.g., Radovan, 2011). We also

excluded cases where scored performances, such as exam scores were combined with self-assessment of learning (e.g., Zhang, 2020).

Rigor and Quality

The deep examination of study features forced by coding for quality did result in some adjustments to the prior data. Three outcomes, two technology and one online learning self-efficacy, from Torun (2020) were collapsed into a single technology self-efficacy outcome using a mini meta-analysis. One external locus of control outcome in Wang and Newlin (2000) was changed to internal locus of control after closer examination of the measure. These adjustments underscore the need for a comprehensive review of primary research quality.

Results

Where relevant, a random-effects (Borenstein et al., 2021) model was used, reflecting a focused (college and vocational) education level coupled with wide variations in study quality, motivational constructs, outcomes, and subject areas. Significance testing was set at the .05 level for all analyses. Finally, a Hunter and Schmidt (1990, 2007) correction was used to adjust for sample size for all analyses and to facilitate corrections for quality as detailed below. To compare moderators, ANOVA style Z-tests were used (Borenstein et al., 2021). Much like a primary study ANOVA, these offer direct comparisons between two groups using a ratio of the difference (r) and square root of the summed variance for any two groups. Similarly, Z-tests were used to determine if effect sizes were significantly different than $r = 0$ or no relationship between motivation and academic performance.

Objective 1: Measurement reporting and study quality

For the first two research questions, effect sizes were adjusted only for sample size. The primary purpose of these questions was to determine which, if any, variations in measurement reporting and methodological quality impact the correlation between motivation as a whole and academic performance.

RQ_1 Measurement reporting quality as a moderator

Most levels (*good*, *attempt*, *none*) in both measurement reliability and measurement validity (see Table 3) were associated with correlations significantly larger than zero. The lone exception was reliability (*attempt* $r = 0.02$). The expectation is that stronger reporting (*good*) should result in larger effect size relationships. For reliability, *good* reporting is associated with reduced measurement error and a closer approximation of the true (larger) effect size. For validity, stronger reporting should more precisely measure the motivational construct and be a closer approximation of the true (larger) effect size.

Table 3

Measurement reporting moderators

Measurement Reporting	*n*	*r*	*CI*[95%]	I^2	*Q*	*p*
Reliability						
good	89	0.14	[0.12, 0.16]	84.2%	557.43	0.01
attempt**	10	0.02	[-0.04, 0.08]	87.2%	70.36	0.53
none*	18	0.19	[0.14, 0.16]	81.1%	89.95	0.01
Validity						
good	34	0.14	[0.11, 0.17]	89.5%	315.29	0.01
attempt**	28	0.10	[0.07, 0.13]	84.5%	174.67	0.01
none	55	0.16	[0.13, 0.18]	77.1%	235.51	0.01
Overall	**117**	**0.14**	**[0.12, 0.15]**	**83.7%**	**711.46**	**0.01**

**Indicates a significant and stronger effect size (r) difference in need of reduction for quality*

***Indicates a significant and weaker effect size (r) difference in need of an increase for quality*

Pairwise comparisons revealed that measurement reliability reported as *none* was significantly higher than *attempt* ($r = 0.19$; $r_{diff} = 0.17$; $Z = 4.06$, $p < 0.01$) suggesting outcomes with no attempt to report measurement reliability may overestimate the effect size (*r*). Two other significant relationships were in an expected direction. Measurement reliability of *attempt* was lower than *good* ($r = 0.02$; $r_{diff} = 0.12$; $Z = 3.72$, $p < 0.01$). For measurement validity, *attempt* was significantly lower than *good* ($r = 0.10$; $r_{diff} = 0.04$; $Z = 2.17$, $p = 0.03$). Both reliability and validity reporting attempts indicate underestimates of the effect size (*r*). Results suggest that variations in measurement reliability and validity reporting do significantly moderate effect sizes (*r*) between motivation and academic performance. Under the assumption that high quality studies are transparent about reporting measurement reliability, outcomes with no reporting or *none* ($n = 18$) were adjusted down for later analyses. Similarly, reliability ($n = 10$) and validity ($n = 22$) reporting *attempts* were adjusted up for later analyses. Overall, these results indicate that measurement reporting quality plays a significant role in moderating the effect size relationships between motivation and academic performance, both over and under-estimating the relationship.

RQ_2 Methodological quality as a moderator

As with measurement reporting, effect sizes were adjusted only for sample size for analysis. These ratings were based on judgements about the methodological quality of each outcome. Out of nine coded MetaQAT plus attrition variables, three indicators of methodological quality acted as significant moderators (see Table 4). All of the significant differences paired *bad/unclear* methods, *bad/unclear* freedom from bias, and *missing* enough data to calculate attrition with stronger relationships.

Table 4

Methodological quality moderators

Methodological Quality	*n*	*r*	*CI*[95%]	I^2	*Q*	*p*
Reliable						
Clear study presentation	111	0.14	[0.12, 0.15]	83.3%	658.86	0.01
bad/unclear presentation	6	0.13	[0.02, 0.24]	90.6%	53.05	0.01
Clear methods & results	92	0.11	[0.10, 0.13]	87.6%	515.12	0.01
bad/unclear methods*	25	0.17	[0.14, 0.19]	69.6%	167.90	0.01
Clear ethics	32	0.12	[0.08, 0.15]	76.7%	133.28	0.01
bad/unclear ethics	85	0.15	[0.13, 0.17]	85.5%	578.75	0.01
Valid						
Methods match study scope	92	0.13	[0.11, 0.15]	84.7%	596.08	0.01
bad/unclear match	25	0.17	[0.13, 0.21]	73.7%	91.29	0.01
Freedom from bias	101	0.13	[0.11, 0.14]	84.4%	640.82	0.01
bad/unclear bias*	16	0.21	[0.17, 0.25]	81.4%	80.83	0.01
Explicit conclusions	24	0.13	[0.11, 0.15]	82.1%	425.55	0.01
bad/unclear conclusions	77	0.14	[0.11, 0.18]	84.3%	146.55	0.01
Generalizable to higher ed	97	0.15	[0.13, 0.16]	85.4%	658.32	0.01
bad/unclear generalizability	20	0.10	[0.05, 0.15]	71.9%	67.71	0.01
Confident in findings	68	0.13	[0.11, 0.15]	85.6%	466.44	0.01
bad/unclear confidence	49	0.15	[0.12, 0.18]	79.6%	235.60	0.01
Attrition reported	61	0.11	[0.09, 0.13]	87.1%	464.22	0.01
Attrition missing*	56	0.17	[0.14, 0.20]	74.7%	217.00	0.01
Overall	**117**	**0.14**	**[0.12, 0.15]**	**83.7%**	**711.46**	**0.01**

**Indicates a significant and stronger effect size (r) difference in need of correction*

For these data, methodological quality clearly mattered. As noted in the preliminary analysis, specific attrition values were calculated when available ($n = 61$). A meta-regression was run to determine if specific attrition rates can predict effect size relationships (r). Results suggested there was no predictive relationship between specific attrition values and academic performance (*beta* = -0.12*)*, $t(55) = 1.15$, $p = 0.25$. There was, however, a marked increase after binary coding attrition and using all available data. When attrition was missing, the effect size went up *(r = 0.17;* $r_{diff} = 0.06$*;* $Z = 3.72$*,* $p < 0.01$*)*. Of the four significant moderators, all of them associated poor methodological quality with increases in the relationship (r) between motivation and academic performance.

Based on findings from the first objective and first two research questions, motivation may be underestimated (reliability and validity reporting *attempts*). More frequently, motivation relationships with academic performance may be over-estimated (attrition, unclear bias, unclear

methods, *no reporting* of reliability). Adjustments are needed to compensate for these significant variations in measurement and methodological quality.

Objective 2: Synchronicity and motivational construct moderators with quality adjusted effect sizes

Building on findings from the first objective, effect sizes were adjusted for sample size as well as methodological quality and measurement reporting quality. For the remaining research questions, overestimates of relationships (*bad/unclear* presentation, ethics, *missing* attrition, *no reporting* of reliability) were adjusted down and underestimates (reliability and validity reporting *attempts*) were adjusted up. The quality adjusted relationship between motivation and academic performance dropped (r_{adj} = 0.10) by almost one third (see Table 5). Note that the overall correlation is not a meaningful value because of the wide range of study characteristics. It is a useful reference point for subgroups and to capture how profound the adjustments for measurement and methodological quality are.

RQ_3 Class synchronicity as moderators, adjusting for quality

Class synchronicity acted as a stark moderator of the relationship (*r*) between motivation and academic performance. Asynchronous class interactions (n = 17, r_{adj} = 0.04) were not significantly different from 0 (Z = 1.73, p = 0.08. In other words, in asynchronous courses, there was no relationship between motivational constructs and academic performance.

Table 5

Class synchronicity moderators

Class Synchronicity	***n***	r_{adj}	***CI*[95%]**	I^2	***Q***
asynchronous***	17	0.04	[-0.01, 0.09]	86.6%	119.51
mixed	48	0.12	[0.09, 0.15]	77.4%	208.24
synchronous	15	0.11	[0.07, 0.15]	94.5%	253.52
unclear	37	0.10	[0.07, 0.12]	80.6%	185.80
Overall (adj)	**117**	**0.10**	**[0.08, 0.12]**	**83.8%**	**717.78**

***Asynchronous classes were not significantly different than r_{adj} = 0.

Asynchronous interactions were also significantly weaker relationships than both *synchronous* (r_{adj} *= 0.04;* r_{diff} = 0.07*; Z = 2.14, p* = 0.03) and *mixed* (r_{diff} = 0.08*; Z = 2.69, p* = 0.01) class interactions. Finally, *unclear* interactions which could have *asynchronous*, *mixed*, and *synchronous* ODDE elements had similar relationships to all forms of synchronicity. Almost a third of the outcomes (n = 37) were *unclear*, lacking enough of a study description to characterize. From these results, it is clear that class synchronicity acted as a significant moderator in the relationship between motivation and academic performance. Truly *asynchronous* classes are rare in this sample (n = 17), and also show no relationship between motivation and academic performance when adjusting for study quality.

RQ$_4$ Motivational constructs as moderators, adjusting for quality

Making corrections for sample size, methodological quality, and measurement reporting quality clearly impacted the strength of relationships between both motivational categories and specific constructs with academic performance. In particular, one entire motivational category, *attributional style* (r_{adj} = 0.04*; Z* = 1.65*, p* = 0.10), and seven specific constructs were not significantly different from r_{adj}= 0 (see Table 6). The results suggest that *internal* and *external locus of control*, r*elative autonomy index*, *extrinsic motivation*, *technology self-efficacy*, *performance avoidance*, and *mastery avoidance* have no relationship with academic performance in ODDE courses.

Table 6

Motivational construct moderators

Motivational Construct	*n*	r_{adj}	*CI*[95%]	I^2	*Q*
Attributional Style* **	14	0.04	[-0.01, 0.09]	71.4%	45.43
internal locus of control***	6	-0.02	[-0.10, 0.06]	87.9%	41.29
external locus of control (rev)***	4	0.10	[-0.02, 0.21]	38.4%	4.87
optimism	4	0.08	[0.04, 0.12]	51.3%	6.16
Self-determination Motivation	28	0.10	[0.06, 0.13]	83.0%	159.20
relative autonomy index***	1	-0.10	[-0.22, 0.02]	-	-
extrinsic motivation***	7	0.04	[-0.04, 0.11]	74.2%	23.22
general motivation	7	0.12	[0.04, 0.21]	88.4%	51.85
intrinsic motivation	13	0.13	[0.08, 0.18]	77.4%	53.05
Self-efficacy	56	0.11	[0.08, 0.13]	83.5%	333.75
technology self-efficacy***	20	0.04	[-0.01, 0.08]	89.8%	186.00
learning self-efficacy	36	0.15	[0.12, 0.18]	76.1%	146.41
Achievement Goals	11	0.10	[0.07, 0.14]	93.0%	143.77
performance avoidance***	2	0.02	[-0.06, 0.09]	76.2%	4.21
mastery avoidance***	3	0.05	[-0.02, 0.13]	95.8%	47.48
mastery approach	4	0.15	[0.09, 0.21]	95.9%	74.05
performance approach	2	0.17	[0.09, 0.25]	96.7%	30.31
Task Value	8	0.16	[0.11, 0.21]	64.5%	19.74
Overall (adj)	**117**	**0.10**	**[0.08, 0.12]**	**83.8%**	**717.78**

****Motivational construct or category was not significantly different than $r_{adj} = 0$*

Note that the relative autonomy index, as a single outcome, does not have heterogeneity (I^2 or Q) statistics. It represents a single extreme outcome. Clearly, along with several other constructs with emerging sets of evidence, additional research is needed. Half of the motivational constructs had four or fewer outcomes. Authors of included studies clearly responded to theoretical calls to vary measurement of self-efficacy by context. Adaptations included scales for online, computer, and even LMS self-efficacy. Aggregated under technology self-efficacy in our

coding, technology self-efficacy was lower than learning self-efficacy *(r_{adj} = 0.04; r_{diff}* = 0.11 *Z = 3.97, p* < 0.01*)*. Neither approached the self-efficacy relationships found among in-person courses (*r* = 0.31 to 0.59) in the meta-analysis of in-person classes replicated here (Richardson et al. 2012). In summary, motivational construct does moderate the relationship between motivation and academic performance after adjusting for study quality. Many constructs are in need of additional research. From available evidence, half of the constructs, representing about one third of all outcomes show no relationship between motivation and academic performance after adjusting for study quality.

Discussion

ODDE scholars displayed a clear interest in the relationship between motivation and academic performance. Ignoring COVID-19 and MOOC studies and focusing on intentionally designed college and vocational courses. A large volume of outcomes (*n* = 117) was represented in this meta-analytic review. That interest increased rapidly in the form of empirical literature with 23 of those outcomes added just over the past year and a half.

On balance, quality analyses suggested that effect sizes (*r*) for online and blended learning are over-estimated by as much as a third. In some cases, there were clear indications of scholarly rigor for measurement reliability reporting. The only two cases of likely underestimates of effect size were both in the area of measurement reporting quality. For *attempts* to report measurement reliability and measurement validity, the effect size was lower than strong reports as expected, and also lower than no reports. This suggests a level of attenuation in need of positive correction. Only 15% (*n* = 18) of outcomes failed to report any measurement reliability and for the vast majority of cases (*n* = 89) reliability specific to that sample was reported. Given the measurement challenges and range of effect sizes for various measures and study contexts already in the literature (e.g., Cho & Summers, 2012; Huang, 2012; Hulleman et al. 2016), this is to be commended and should continue. Measurement validity reporting was far less clear with almost half (*n* = 55) of the outcomes lacking any kind of report. In many ways, validity is even more crucial given reported relationships between factors and effect size (Huang, 2012). Clearly ODDE scholars made attempts at following theoretical recommendations to assess self-efficacy in domain-specific ways (Bandura, 1986, 2006). However, it is also clear that variations on technology self-efficacy may not align with Bandura's thinking about domain. In a broad sense, through the lens of motivational constructs at least, it is clear that ODDE is a different measurement context for motivation than in-person settings.

Under-reporting was also present in methodological quality. Attrition had so many missing values (*n* = 56) that it lacked sensitivity to use specific attrition rates despite a wide range (see Appendix A) and had to be binary coded to adjust for quality. Methodological quality was not a significant factor most of the time, but when it was significant, lower quality (bad/unclear methods, bad/unclear bias, and no reports of attrition) resulted in significantly stronger correlation values that required correction. Inflated effect sizes for less rigorous methods matched prior reviews of in-person and digital learning (Macnamara et al., 2023). Contextualizing quality in terms of MetaQAT, casting a wide net for relevancy was rewarded

with quality challenges to methodological reliability and methodological validity. The ODDE literature for relationships between motivation and academic performance did appear to have questionable applicability. Note that strong MetaQAT indicators should be more sensitive to real differences in outcomes and more likely to show higher correlations, which was never the case.

Across the combination of measurement reporting and methodological quality, significant differences existed. Quantifying and adjusting for those differences reduced the overall effect size from $r = 0.14$ to $r_{adj} = 0.10$. Quality adjustments similarly resulted in significant changes when exploring mean difference (*g*) variations in interaction strength for online and blended learners (Bernard et al., 2009). The work of comprehensively quantifying and adjusting for quality differences is a direct benefit of this review, and consistent with prior reviews (Macnamara et al., 2023) of in-person learners. Based on the number of quality moderators and the magnitude of effect size reduction from quality adjustments, researchers should strive for more rigorous measurement validity, keep up efforts on measurement reliability, and recognize that reductions in methodological quality may overestimate relationships between motivation and academic performance. An indirect benefit of intense quality coding was the important adjustments to motivational construct assignments for two different studies. Thus, for reviewers, both coding and adjusting for quality were clearly important for properly characterizing and contextualizing the ODDE literature as it intersects with motivation. In a dramatic shift from in-person meta-analyses, several moderators such as asynchronous class interactions and half of the motivational constructs were not just weak effect sizes, they were so weak that they supported the null hypothesis of no relationship between motivation and academic performance in ODDE. Bernard et al. (2009) did not find overall mean differences (*g*) between asynchronous, synchronous or a mix of synchronicity when examined in isolation, but they did detect differences when simultaneously accounting for low, moderate, and high interaction. Our results showed that synchronicity did matter across all observations, specifically that asynchronous settings showed no relationship ($r_{adj} = 0.04$) between motivation and academic performance. As a key finding, our results were a sharp contrast from the widespread agreement about motivation (Richardson et al., 2012), including specific recommendations to enhance academic self-efficacy in order to improve academic outcomes in classroom settings (Honicke et al., 2020). Specific to the ODDE literature, intentional choices to select for (Krebs-Lazendic et al., 2026) or code for synchronous experiences in reviews is warranted—at least in the context of motivation.

Limitations and Future Work

Coding for quality relied on the transparency of primary research. There was an ordinal relationship between good validity (actual analysis as part of a given study) an attempt (citing the validity efforts of another study) or no reporting. Caution is warranted when generalizing findings, especially for validity given the large number ($n = 55$) of studies with no reporting. Additionally, both validity and reliability at times reported specific scores or analyses. In the case of measurement reliability, the vast majority of outcomes reported coefficients for their specific sample ($n = 89$). In short—there was hidden precision in measurement validity and reliability that should be explored. These actual values could be used to model unattenuated correlations (Hunter & Schmidt, 2007) by means of a psychometric meta-analysis, a logical next

step for this research. That work would not be possible without the existing quality efforts of authors in this area of literature—particularly for measurement reliability.

One moderator, applicability to higher education showed weak inter-rater reliability. While it was not a significant moderator of effect size (*r*), measurement error could have masked meaningful differences between good and bad or unclear applicability. Refinement of MetaQAT may be warranted, particularly with uses like this meta-analysis that represent literatures outside of the public health context it was created for (Rosella et al. 2016). Note that this limitation was ameliorated given the consensus coding context for the reported results.

The evidence base for this literature clearly has missing data, such as measurement validity that held true for prior examinations of the ODDE literature. Bernard et al. (2004) reported as much as 60% of study features were missing. Missing data in this study have clearly gone down, suggesting the ODDE literature is improving. The rate does still warrant caution when generalizing to the literature as a whole, especially because both bad/unclear quality elements at times and not reporting attrition were significant moderators of effect size (r). Relevance to select, keep, and code for the full range of study and measurement quality available was an intentional choice for this analysis, in part to quantify the impact and to promote better reporting of future ODDE research.

While all of the motivational constructs do have ties to social cognitive theory, they also have rich independent underlying motivational theories. This was not the first attempt to synthesize across an eclectic mix of motivational constructs and their underlying theories (Lazowski & Hulleman, 2016). Future reviewers may want to specifically code for motivational theory and measurement alignment, or incorporate additional orientations such as student interaction (Moore, 1989) extending the work of Bernard et al., 2009, community of inquiry (Shea et al. 2022), or construct validity, extending the work of Hulleman et al., 2010 for ODDE learners and in a more comprehensive way than their focus on achievement goal theory.

Conclusion

This review combines prior efforts to examine measurement reliability (Richardson et al., 2012), validity (Huang et al. 2012), or methodological quality (Macnamara et al., 2023). Unique to this review space, we use an established quality tool (Rosella et al. 2016). Identifying and adjusting for significant quality moderators we contribute a much needed update of synchronicity bridging more than two decades of research (Bernard et al., 2009) as well as providing a quality adjusted relationships for a range of motivational constructs (Richardson et al., 2012; Walker et al., 2024).

For relationships between motivation and academic performance it was clear that quality did indeed matter. Initially, the primary importance looks academic. ODDE authors are to be commended on their attention and reporting of measurement reliability. For measurement quality reporting as a whole, effect size relationships were largely expected or indicated possible under-estimates of motivation and academic performance. There is a clear need to expand reporting of validity, especially because of the challenges in the online measurement space and how validity

clearly differed from in-person contexts. It is possible that ODDE motivation may need to be measured as part of a larger set of constructs, like the representations of self-efficacy alongside self-regulated learning in the Community of Inquiry framework (Shea et al. 2022) rather than in isolation. The scholarly community may also want to stop exploring variations of technology self-efficacy as an attempt to address calls for domain specificity (Bandura, 2006). Reviewers should clearly address quality in a comprehensive way, not only because of these findings but because of past reviews (Bernard et al., 2009; L.-F. Huang et al., 2012; Macnamara et al., 2023).

Modest correlations for motivation and academic performance (Walker et al., 2024) are even lower when adjusted for measurement reporting quality and methodological quality, to the point that whole categories and half of the constructs show no relationship at all. These findings underscore the need to replicate prior efforts to provide context through exploring study quality (Bernard et al., 2004) or make adjustments for measurement (Richardson et al., 2012). Policy makers should be cautious about extending in-person relationships and recommendations (Honicke et al. 2020) to online learning and should also pay close attention to the impact of truly asynchronous class experiences. These results add to the extant findings that ODDE is a different setting (Cho & Summers, 2012) in regards to motivation. Research quality, review implications, and main findings for synchronicity and motivation should be of interest to policymakers. Policy makers and practitioners may want to explore alternatives or supplements to asynchronous classes in this context. This is an applied field, and there is much to be learned. If motivation does not carry the kind of relationships in ODDE settings that it does face to face—we need to explore why. It could be due to the measurement context; it could be a function of factors for non-traditional students that trump motivation. Quality research is best done with appropriate funding, time, and as a partnership between designers, instructors and researchers—as well as the many academics the engage in multiple roles simultaneously. Given recent increases in online and blended learning experiences (Bay, 2026), the interest in student success models (Jacobsen, 2025), and the high stakes of a declining higher education student population (National Center for Educational Statistics, 2024), practitioners and policy makers stand to benefit from understanding why motivation has such a different relationship for ODDE learners.

References

Alkış, N., & Temizel, T. T. (2018). The impact of motivation and personality on academic performance in online and blended learning environments. *Journal of Educational Technology & Society*, *21*(3), 35–47. https://doi.org/10.30191/ETS.201807_21(3).0004

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, *80*(3), 260–267. https://doi.org/10.1037/0022-0663.80.3.260

Bahçekapili, E., & Karaman, S. (2020). A path analysis of five-factor personality traits, self-efficacy, academic locus of control and academic achievement among online students. *Knowledge Management & E-Learning*, *12*(2), 191–208. https://doi.org/10.34105/j.kmel.2020.12.010

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc.

Bandura, A. (2006). Guide for constructing self-efficacy scales. in f. pajares & t. urdan (eds.), *Self-efficacy beliefs of adolescents* (Vol. 5, pp. 307–337). Information Age Publishing.

Bardach, L., Oczlon, S., Pietschnig, J., & Lüftenegger, M. (2020). Has achievement goal theory been right? a meta-analysis of the relation between goal structures and personal achievement goals. *Journal of Educational Psychology*, *112*(6), 1197–1220. https://doi.org/10.1037/edu0000419

Bawa, P. (2016). Retention in online courses: Exploring issues and solutions—a literature review. *SAGE Open*, *6*(1). https://doi.org/10.1177/2158244015621777

Bay, J. (2026, January 15). *Distance Education Enrollment Surges Nationwide*. Inside Higher Education. https://www.insidehighered.com/news/student-success/academic-life/2026/01/15/distance-education-enrollment-surges-nationwide

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, *79*(3), 1243–1289. https://doi.org/10.3102/0034654309333844

Bernard, R. M., Abrami, P. C., Wade, A., Borokhovski, E., & Lou, Y. (2004, October 19–23). The effects of synchronous and asynchronous distance education: A meta-analytical assessment of Simonson's "equivalency theory" [Conference paper]. 27th Annual Convention of the Association for Educational Communications and Technology, Chicago, IL, United States. https://eric.ed.gov/?id=ED485078

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2021). *Introduction to Meta-Analysis* (2nd ed.). Wiley.

Chang, C.-C., Yen, W.-H., Liu, S.-C., & Kuo, C.-T. (2026). Learning presence in the revised community of inquiry framework: A systematic review of empirical research (2010–2024). *The Internet and Higher Education*, *68*, 101058. https://doi.org/10.1016/j.iheduc.2025.101058

Cho, M. H., & Summers, J. (2012). Factor validity of the motivated strategies for learning questionnaire (MSLQ) in asynchronous online learning environments. *Journal of Interactive Learning Research*, *23*(1), 5-28.

Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: Ii. resolving the paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 551–558. https://doi.org/10.1016/0895-4356(90)90159-m

de Araujo, J., Gomes, C. M. A., & Jelihovschi, E. G. (2023). The factor structure of the motivated strategies for learning questionnaire (MSLQ): New methodological approaches and evidence. *Psicologia, Reflexao e Critica: Revista Semestral Do Departamento de Psicologia Da UFRGS*, *36*(1), 38. https://doi.org/10.1186/s41155-023-00280-0

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*(3), 501–519. https://doi.org/10.1037/0022-3514.80.3.501

Fromm, Y. M., Martin, F., Gezer, T., & Ifenthaler, D. (2025). Best practices for conducting systematic reviews: Perspectives of experienced systematic review researchers in educational sciences. *Technology, Knowledge and Learning*, *30*(1), 1–28. https://doi.org/10.1007/s10758-025-09819-9

Gardner, A., Ibrahim, M., Karamarkovich, S., Holsapple, M., & Shapiro, D. (2024). PDP insights: Early indicators of timely completers. in *National Student Clearinghouse*. National Student Clearinghouse Research Center. https://eric.ed.gov/?id=ED673395

Goulão, M. (2014). The relationship between self-efficacy and academic achievement in adults' learners. *Athens Journal of Education*, *1*, 237–246. https://doi.org/10.30958/aje.1-3-4

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Hilpert, J. C., Stempien, J., van der Hoeven Kraft, K. J., & Husman, J. (2013). Evidence for the latent factor structure of the MSLQ: A new conceptualization of an established questionnaire. *SAGE Open*, *3*(4), 2158244013510305. https://doi.org/10.1177/2158244013510305

Hodges, C. B. (2008). Self-efficacy, motivational email, and achievement in an asynchronous math course. *The Journal of Computers in Mathematics and Science Teaching*, *27*(3), 265–285.

Honicke, T., Broadbent, J., & Fuller-Tyszkiewicz, M. (2020). Learner self-efficacy, goal orientation, and academic achievement: Exploring mediating and moderating relationships. *Higher Education Research & Development*, *39*(4), 689–703. https://doi.org/10.1080/07294360.2019.1685941

Howard, J. L., Bureau, J., Guay, F., Chong, J. X. Y., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *16*(6), 1300–1323. https://doi.org/10.1177/1745691620966789

Huang, C. (2012). Discriminant and criterion-related validity of achievement goals in predicting academic achievement: A meta-analysis. *Journal of Educational Psychology*, *104*(1), 48–73. https://doi.org/10.1037/a0026223

Huang, L.-F., Wu, T.-Y., Hou, S.-H., & Horng, R.-Y. (2012). Effects of feedback type, problem presentation, and practice on problem solving. *Bulletin of Educational Psychology*, *43*(4), 855–874.

Hulleman, C.S., Barron, K.E., Kosovich, J.J., Lazowski, R.A. (2016). Student Motivation: Current Theories, Constructs, and Interventions Within an Expectancy-Value Framework. In: Lipnevich, A., Preckel, F., Roberts, R. (eds) Psychosocial Skills and School Systems in the 21st Century. The Springer Series on Human Exceptionality. Springer, Cham. https://doi.org/10.1007/978-3-319-28606-8_10

Hulleman, C. S., Schrager, S. M., Bodmann, S. M., & Harackiewicz, J. M. (2010). A Meta-Analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin*, *136*(3), 422–449. https://doi.org/10.1037/a0018947

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting Error and Bias in Research Findings*. Sage Publications.

Hunter, J. E., & Schmidt, F. L. (2007). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. SAGE Publications, Inc.

Jacobsen, L. J. (2025). *ACCESS: A Framework for Student Success*. Inside Higher Ed. https://www.insidehighered.com/opinion/views/2025/08/06/framework-organizing-student-success-efforts-opinion

Johnson, N., Seaman, J., & Poulin, R. (2022). Defining different modes of learning: Resolving confusion and contention through consensus. *Online Learning*, *26*(3). https://doi.org/10.24059/olj.v26i3.3565

Joo, Y. J., Lim, K. Y., & Kim, J. (2013). Locus of control, self-efficacy, and task value as predictors of learning outcome in an online university context. *Computers & Education*, *62*, 149–158. https://doi.org/10.1016/j.compedu.2012.10.027

Katrak, P., Bialocerkowski, A. E., Massy-Westropp, N., Kumar, S., & Grimmer, K. A. (2004). A systematic review of the content of critical appraisal tools. *BMC Medical Research Methodology*, *4*, 22. https://doi.org/10.1186/1471-2288-4-22

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

Kim, J.-Y. (2012). A study on learners' perceptional typology and relationships among the learner's types, characteristics, and academic achievement in a blended E-Education environment. *Computers & Education*, *59*(2), 304–315.

Krebs-Lazendic, L., Dello Stritto, M. E., & Park, E. (2026). A critical review of research on synchronous instructor presence in online learning: Elusive evidence and mixed results. *Quarterly Review of Distance Education*, 27(1)*,* 1–19. https://doi.org/10.1108/QRDE-09-2025-0016

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, *30*(3), 411–433. https://doi.org/10.1111/j.1468-2958.2004.tb00738.x

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation interventions in education: A Meta-Analytic review. *Review of Educational Research*, *86*(2), 602–640. https://doi.org/10.3102/0034654315617832

Lee, K. (2017). Rethinking the ACCESSibility of online higher education: A historical review. *The Internet and Higher Education*, *33*, 15–23. https://doi.org/10.1016/j.iheduc.2017.01.001

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Sage.

Liu, C., Sands-Meyer, S., & Audran, J. (2019). The effectiveness of the student response system (SRS) in english grammar learning in a flipped english as a foreign language (EFL) class. *Interactive Learning Environments*, *27*(8), 1178–1191.

Lucas, M., & Vicente, P. N. (2023). A double-edged sword: Teachers' perceptions of the benefits and challenges of online teaching and learning in higher education. *Education and Information Technologies*, *28*(5), 5083–5103. https://doi.org/10.1007/s10639-022-11363-3

Macnamara, B. N., & Burgoyne, A. P. (2023). Do growth mindset interventions impact students' academic achievement? a systematic review and meta-analysis with recommendations for best practices. *Psychological Bulletin*, *149*(3–4), 133–173. https://doi.org/10.1037/bul0000352

Mahmud, M., Freeman, B., & Abu Bakar, M. (2022). Technology in education: Efficacies and outcomes of different delivery methods. *Interactive Technology and Smart Education*, *19*(1), 20–38. https://doi.org/10.1108/ITSE-01-2021-0021

Martin, F., Wu, T., Wan, L., & Xie, K. (2022). A Meta-Analysis on the Community of Inquiry Presences and Learning Outcomes in Online and Blended Learning Environments. *Online Learning (Newburyport, Mass.)*, *26*(1), 325. https://doi.org/10.24059/olj.v26i1.2604

Martin, F., Tutty, J. I., & Su, Y. (2010). Influence of learning management systems self-efficacy on e-learning performance. *Journal on School Educational Technology*, *5*(3), 26–35.

Marzi, G., Balzano, M., & Marchiori, D. (2024). K-Alpha calculator–Krippendorff's alpha calculator: A user-friendly tool for computing Krippendorff's alpha inter-rater reliability coefficient. *MethodsX*, *12*, 102545. https://doi.org/10.1016/j.mex.2023.102545

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

McPhaul-Moore, E. (2013). Predictors of online learning success in a rural community college [doctoral dissertation, northcentral university]. in *ProQuest LLC*. https://www.proquest.com/docview/1444627614

Moore, M. G. (1989). Three types of interaction. *American Journal of Distance Education*, *3*(2), 1–7. https://doi.org/10.1080/08923648909526659

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A Meta-Analytic investigation. *Journal of Counseling Psychology*, *38*(1), 30–38. https://doi.org/10.1037/0022-0167.38.1.30

National Center for Education Statistics. (2024). *College Enrollment Rates* (Condition of Education). Department of Education, Institute of Education Sciences. https://nces.ed.gov/programs/coe/indicator/cpb

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic reviews*, *5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinnrdd, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *71*. https://doi.org/10.1136/bmj.n71

Perkowski, J. (2013). The role of gender in distance learning: A Meta-Analytic review of gender differences in academic performance and self-efficacy in distance learning. *Journal of Educational Technology Systems*, *41*(3), 267–278. https://doi.org/10.2190/ET.41.3.e

Pintrich, P. R. (1991). *A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ)*. https://eric.ed.gov/?id=ED338122

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813. https://doi.org/10.1177/0013164493053003024

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353–387. https://doi.org/10.1037/a0026838

Rosella, L., Bowman, C., Pach, B., Morgan, S., Fitzpatrick, T., & Goel, V. (2016). The development and validation of a meta-tool for quality appraisal of public health evidence: Meta quality appraisal tool (MetaQAT). *Public Health*, *136*, 57–65. https://doi.org/10.1016/j.puhe.2015.10.027

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Savoji, A. P. (2013). Motivational strategies and academic achievement in traditional and virtual university students. *Procedia - Social and Behavioral Sciences*, *84*, 1015–1020. https://doi.org/10.1016/j.sbspro.2013.06.691

Senko, C., & Dawson, B. (2017). Performance-approach goal effects depend on how they are defined: Meta-Analytic evidence from multiple educational outcomes. *Journal of Educational Psychology*, *109*(4), 574–598. https://doi.org/10.1037/edu0000160

Shea, B. J., Grimshaw, J. M., Wells, G. A., Boers, M., Andersson, N., Hamel, C., Porter, A. C., Tugwell, P., Moher, D., & Bouter, L. M. (2007). Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. *BMC Medical Research Methodology*, *7*, 10. https://doi.org/10.1186/1471-2288-7-10

Shea, P., Richardson, J., & Swan, K. (2022). Building bridges to advance the community of inquiry framework for online learning. *Educational Psychologist*, *57*(3), 148–161. https://doi.org/10.1080/00461520.2022.2089989

Stephen, J. S., Rockinson-Szapkiw, A. J., & Dubay, C. (2020). Persistence model of non-traditional online learners: Self-efficacy, self-regulation, and self-direction. *American Journal of Distance Education*, *34*(4), 306–321. https://doi.org/10.1080/08923647.2020.1745619

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, *366*, l4898. https://doi.org/10.1136/bmj.l4898

Storrings, D. (2005). Attrition in distance education: A meta-analysis. *Instructional Design, Development and Evaluation - Dissertations*. https://surface.syr.edu/idde_etd/8

Theobald, M. (2021). Self-regulated learning training programs enhance university students' academic performance, self-regulated learning strategies, and motivation: A meta-analysis. *Contemporary Educational Psychology*, *66*, 101976. https://doi.org/10.1016/j.cedpsych.2021.101976

Tladi, L. S. (2017). Perceived ability and success: Which self-efficacy measures matter? a distance learning perspective. *Open Learning: The Journal of Open, Distance and e-Learning*, *32*(3), 243–261. https://doi.org/10.1080/02680513.2017.1356711

Torun, E. D. (2020). Online distance learning in higher education: E-learning readiness as a predictor of academic achievement. *Open Praxis*, *12*(2), 191–208. https://search.informit.org/doi/10.3316/informit.352722164544178

Walker, A., Aguiar, N. R., Soicher, R., Kuo, Y.-C., & Resig, J. (2024). Exploring the relationships between motivation and academic performance among online and blended learners: A Meta-Analytic review. *Online Learning*, *28*(4). https://doi.org/10.24059/olj.v28i4.4602

Wang, A. Y., & Newlin, M. H. (2000). Characteristics of students who enroll and succeed in psychology web-based classes. *Journal of Educational Psychology*, *92*(1), 137–143. https://doi.org/10.1037/0022-0663.92.1.137

Wang, L. (2022). Student Intrinsic Motivation for Online Creative Idea Generation: Mediating Effects of Student Online Learning Engagement and Moderating Effects of Teacher Emotional Support. *Frontiers in Psychology*, *13*, 954216. https://doi.org/10.3389/fpsyg.2022.954216

Warren, L., Reilly, D., Herdan, A., & Lin, Y. (2020). Self-efficacy, performance and the role of blended learning. *Journal of Applied Research in Higher Education*, *13*(1), 98–111. https://doi.org/10.1108/JARHE-08-2019-0210

Weiner, B. (2010). The development of an attribution-based theory of motivation: A history of ideas. *Educational Psychologist*, *45*(1), 28–36. https://doi.org/10.1080/00461520903433596

West, R. E., & Martin, F. (2024). What type of paper are you writing? a taxonomy of review and theory scholarship distinguished by their summary and advocacy arguments. *Educational Technology Research and Development*, *72*(5), 2443–2476. https://doi.org/10.1007/s11423-023-10233-0

Whiting, P., Savović, J., Higgins, J. P. T., Caldwell, D. M., Reeves, B. C., Shea, B., Davies, P., Kleijnen, J., & Churchill, R. (2016). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, *69*, 225–234. https://doi.org/10.1016/j.jclinepi.2015.06.005

Zhao, Y., Lei, J., Lai, B. Y. C., & Tan, H. S. (2005). What makes the difference? a practical analysis of research on the effectiveness of distance education. *Teachers College Record*, *107*(8), 1836–1884. https://doi.org/10.1111/j.1467-9620.2005.00544.x

Acknowledgements

This work was supported by a three-year seminar series on meta-analysis by Oregon State University. Authors received travel support and Andrew Walker received a small stipend as co-leader of the seminar.

ONLINE LEARNING
Volume 30, Issue SI2 (2026) · e5822
https://doi.org/10.24059/olj.5822

Appendix A -- Studies, Outcomes, and Coding

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
Al Fadda (2019)														
learning SE	asynch	attempt	good	--	--	--	--	--	--	--	--	80%	0.29	0.21
Alkis & Temizel (2018)														
intrinsic motiv	mixed	attempt	good	good	good	good	good	--	--	good	--	52%	0.17	0.12
learning SE	mixed	attempt	good	good	good	good	good	--	--	good	--	52%	0.22	0.17
intrinsic motiv	mixed	attempt	good	good	good	good	good	--	--	good	--	52%	0.44	0.39
learning SE	mixed	attempt	good	good	good	good	good	--	--	good	--	52%	0.03	-0.02
Amin (2019)														
intrinsic motiv	mixed	none	none	good	--	--	--	good	good	good	--	--	0.65	0.54
extrinsic motiv	mixed	none	none	good	--	--	--	good	good	good	--	--	0.48	0.37
learning SE	mixed	none	none	good	--	--	--	good	good	good	--	--	0.57	0.46
Bahçekapili & Karaman (2020)														
internal LoC	synch	none	good	good	good	--	good	--	--	good	good	18%	0.03	-0.04
external LoC	synch	none	good	good	good	--	good	--	--	good	good	18%	0.16	0.09
learning SE	synch	none	good	good	--	--	good	--	--	good	good	18%	0.14	0.04
Baturay & Yukselturk (2015)														
technology SE	mixed	none	good	good	good	--	good	good	--	--	--	67%	0.11	0.11
Broadbent & Howe (2023)														
technology SE	mixed	none	good	good	--	good	good	good	good	good	--	--	0.27	0.21
Cannon et al. (2021)														
learning SE	mixed	good	good	good	good	good	good	good	good	good	good	25%	0.05	0.05
Chang & Ho (2009)														
internal LoC	asynch	none	good	good	good	--	--	good	good	good	--	--	0.17	0.14
Chang et al. (2014)														
technology SE	unclear	none	good	good	good	--	good	--	--	good	good	8%	0.24	0.17

Corresponding author: Dr. Andrew Walker, andy.walker@usu.edu
This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 The Author(s).

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
Study/Construct	**Delivery**	**Validity[a]**	**Reliability[b]**	**1[c]**	**2[d]**	**3[e]**	**4[f]**	**5[g]**	**6[h]**	**7[i]**	**8[j]**	**9[k]**	***r***	r_{adj}
Chen & Jang (2010)														
RAI	mixed	good	good	good	--	good	good	good	good	good	good	--	-0.04	-0.10
Cheng et al. (2023)														
learning SE	unclear	attempt	attempt	good	--	good	good	good	good	good	good	--	0.10	0.18
task value	unclear	attempt	attempt	good	--	good	good	good	good	good	good	--	0.07	0.15
Cho & Shen (2013)														
intrinsic motiv	asynch	none	good	good	--	good	good	good	good	good	--	--	-0.09	-0.15
extrinsic motiv	asynch	none	good	good	--	good	good	good	good	good	--	--	-0.08	-0.14
learning SE	asynch	none	good	good	--	good	good	good	good	good	--	--	0.18	0.12
Cigdem & Ozturk (2016)														
technology SE	mixed	attempt	good	good	--	--	good	--	--	--	--	--	0.21	0.10
general motiv	mixed	attempt	good	good	--	--	good	--	--	--	--	--	0.16	0.05
de Fátima Goulão (2014)														
learning SE	unclear	none	good	good	--	--	good	--	good	good	good	--	0.29	0.16
DeTure (2004)														
technology SE	mixed	good	good	good	--	--	good	good	good	good	good	55%	-0.03	-0.06
Ergul (2004)														
learning SE	unclear	none	good	good	--	--	good	good	good	good	good	--	0.24	0.18
mastery approach	unclear	none	good	good	--	--	good	good	good	good	good	--	0.01	-0.05
Gerlich et al. (2009)														
external LoC	asynch	none	none	good	--	--	good	good	good	--	--	--	0.08	-0.03
Goad et al. (2021)														
mastery approach	asynch	attempt	none	good	--	good	--	--	good	good	--	5%	0.17	0.04
Gravelle et al. (2023)														

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
learning SE	mixed	none	good	good	good	good	good	good	good	good	good	12%	0.08	0.08
learning SE	mixed	none	good	good	--	good	good	good	good	good	good	24%	0.10	0.07

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
Gravelle et al. (2023)														
learning SE	mixed	none	good	good	good	good	good	good	good	good	good	12%	0.08	0.08
learning SE	mixed	none	good	good	--	good	good	good	good	good	good	24%	0.10	0.07
Hamm et al. (2019)														
optimism	unclear	good	good	good	good	--	good	good	good	good	good	19%	0.07	0.07
optimism	unclear	good	good	good	good	--	good	good	good	good	good	2%	0.04	0.04
optimism	unclear	good	good	good	good	--	good	good	good	good	good	19%	0.17	0.17
optimism	unclear	good	good	good	good	--	good	good	good	good	good	2%	0.04	0.04
Hefter & Nitsch (2024)														
mastery avoidance	synch	none	good	good	--	good	good	good	good	good	--	--	-0.21	-0.27
learning SE	synch	none	good	good	--	good	good	good	good	good	--	--	0.04	-0.02
learning SE	synch	none	good	good	--	good	good	good	good	good	--	--	0.30	0.24
Hobson & Puruhito (2018)														
learning SE	unclear	attempt	attempt	good	good	--	good	--	good	good	good	28%	0.31	0.38
Hodges (2008)														
learning SE	mixed	attempt	good	--	good	good	good	--	good	good	--	56%	0.28	0.23
Jadric, Bubas & Hutinski (2010)														
technology SE	mixed	attempt	good	good	--	--	good	good	good	--	good	38%	0.15	0.14
Jiang et al. (2024)														
general motiv	mixed	good	good	good	--	good	--	good	good	good	good	--	0.35	0.29
Johnson et al. (2008)														
technology SE	asynch	good	good	good	--	--	good	good	good	good	good	--	0.27	0.21
Joo, Lim, & Kim (2013)														

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
internal LoC	asynch	good	good	good	good	--	good	good	good	good	good	8%	0.07	0.07
learning SE	asynch	good	good	good	good	--	good	good	good	good	good	8%	0.22	0.22
task value	asynch	good	good	good	good	--	good	good	good	good	good	8%	0.25	0.25
task value	asynch	good	good	good	good	--	good	good	good	good	good	8%	0.19	0.19
Study/Construct	**Delivery**	**Validity[a]**	**Reliability[b]**	**1[c]**	**2[d]**	**3[e]**	**4[f]**	**5[g]**	**6[h]**	**7[i]**	**8[j]**	**9[k]**	***r***	r_{adj}
Kim (2012)														
learning SE	mixed	none	none	good	good	--	--	good	good	good	--	--	0.40	0.32
learning SE	mixed	none	none	good	good	--	--	good	good	good	--	--	0.40	0.32
Kong et al. (2025)														
intrinsic motiv	asynch	none	good	good	--	--	good	good	good	--	--	--	0.16	0.10
extrinsic motiv	asynch	none	good	good	--	--	good	good	good	--	--	--	0.16	0.10
Lee & Choi (2013)														
internal LoC	unclear	good	good	good	good	--	good	good	good	good	good	--	0.29	0.26
Lin et al. (2024)														
intrinsic motiv	mixed	none	good	good	--	--	good	good	good	good	good	--	0.34	0.28
extrinsic motiv	mixed	none	good	good	--	--	good	good	good	good	good	--	0.06	0.00
task value	mixed	none	good	good	--	--	good	good	good	good	good	--	0.18	0.12
learning SE	mixed	none	good	good	--	--	good	good	good	good	good	--	0.28	0.22
intrinsic motiv	mixed	none	good	good	--	--	good	good	good	good	good	--	0.37	0.31
extrinsic motiv	mixed	none	good	good	--	--	good	good	good	good	good	--	0.04	-0.02
task value	mixed	none	good	good	--	--	good	good	good	good	good	--	0.40	0.34
learning SE	mixed	none	good	good	--	--	good	good	good	good	good	--	0.26	0.20
Liu et al. (2019)														
learning SE	mixed	none	none	good	good	--	good	--	good	--	--	--	-0.06	-0.21
general motiv	mixed	none	none	good	good	--	good	--	good	--	--	--	-0.58	-0.53
Liu, Ma, & Chen (2024)														
extrinsic motiv	unclear	good	good	good	good	--	good	good	good	good	good	8%	-0.10	-0.10
intrinsic motiv	unclear	good	good	good	good	--	good	good	good	good	good	8%	0.08	0.08

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
Lynch & Dembo (2004)														
learning SE	mixed	good	good	good	good	--	good	good	good	good	good	73%	0.29	0.29
technology SE	mixed	good	good	good	good	--	good	good	good	good	good	73%	-0.09	-0.09
Martin et al. (2010)														
technology SE	mixed	good	good	good	--	--	good	good	good	good	good	--	0.20	0.14
Study/Construct	**Delivery**	**Validity[a]**	**Reliability[b]**	**1[c]**	**2[d]**	**3[e]**	**4[f]**	**5[g]**	**6[h]**	**7[i]**	**8[j]**	**9[k]**	***r***	r_{adj}
McPhaul-Moore (2013)														
learning SE	unclear	attempt	none	good	good	good	good	good	good	good	good	24%	0.23	0.20
intrinsic motiv	unclear	attempt	none	good	good	good	--	good	good	good	--	29%	0.09	0.06
Morris (2005)														
internal LoC	unclear	attempt	attempt	good	--	good	--	good	good	good	--	87%	-0.26	-0.15
Neroni et al. (2022)														
learning SE	asynch	none	good	good	good	good	good	good	--	good	--	77%	0.05	0.05
Noteborn et al. (2012)														
task value	unclear	none	good	good	--	--	good	good	good	good	good	--	0.14	0.08
Ongor & Uslusoy (2023)														
general motiv	synch	good	good	good	good	good	good	--	good	good	good	--	0.23	0.13
Parker et al. (2021)														
intrinsic motiv	unclear	none	good	good	--	--	good	good	good	good	good	--	0.29	0.23
task value	unclear	none	good	good	--	--	good	good	good	good	good	--	0.07	0.01
Redstone (2023)														
technology SE	asynch	attempt	good	good	--	good	good	good	good	good	good	55%	-0.10	-0.11
Remedios & Richardson (2012)														
mastery approach	unclear	good	good	good	good	--	good	good	good	good	good	35%	0.00	0.00
mastery avoidance	unclear	good	good	good	good	--	good	good	good	good	good	35%	-0.01	-0.01
performance approach	unclear	good	good	good	good	--	good	good	good	good	good	35%	-0.01	-0.01

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
performance avoidance	unclear	good	good	good	good	--	good	good	good	good	good	35%	0.08	0.08
Runyon (2013)														
technology SE	unclear	good	attempt	good	good	good	good	good	good	good	good	58%	0.28	0.40
Sankaran & Bui (2001)														
general motiv	unclear	none	good	--	--	--	--	good	good	good	--	--	0.58	0.52
Study/Construct	**Delivery**	**Validity[a]**	**Reliability[b]**	**1[c]**	**2[d]**	**3[e]**	**4[f]**	**5[g]**	**6[h]**	**7[j]**	**8[k]**	**9[l]**	***r***	r_{adj}
Savoji (2013)														
intrinsic motiv	unclear	attempt	good	good	good	--	good	good	good	good	good	--	0.06	0.05
extrinsic motiv	unclear	attempt	good	good	good	--	good	good	good	good	good	--	0.06	0.05
task value	unclear	attempt	good	good	good	--	good	good	good	good	good	--	0.13	0.12
learning SE	unclear	attempt	good	good	good	--	good	good	good	good	good	--	0.29	0.28
Simmering, Posey & Piccoli (2009)														
technology SE	mixed	attempt	good	good	good	--	good	good	good	good	good	12%	0.29	0.31
general motiv	mixed	attempt	good	good	good	--	good	good	good	good	good	12%	-0.07	-0.05
Stephen et al. (2020)														
technology SE	unclear	attempt	attempt	--	good	--	good	good	good	good	good	54%	-0.78	-0.53
learning SE	unclear	attempt	attempt	good	good	--	good	good	good	good	good	54%	-0.03	0.11
Tai (2016)														
learning SE	mixed	good	good	good	good	good	good	good	good	good	good	--	-0.29	-0.32
Talsma et al. (2023)														
learning SE	mixed	none	none	good	--	good	--	good	good	good	--	--	0.31	0.20
learning SE	mixed	none	good	good	--	good	good	good	good	good	--	--	0.14	0.08
Tladi (2017)														
learning SE	synch	attempt	good	good	good	--	good	good	good	good	good	98%	0.13	0.15
technology SE	synch	attempt	good	good	good	--	good	good	good	good	good	98%	-0.74	-0.53
Torun (2020)														
general motiv	synch	good	good	good	good	--	--	--	--	good	--	1%	0.51	0.44

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
technology SE	synch	good	good	good	good	--	good	--	--	good	--	1%	0.16	0.09
Tseng, Chen, & Lin (2023)														
learning SE	mixed	attempt	good	good	--	good	good	--	good	good	--	--	0.18	0.07
Wadsworth et al. (2007)														
learning SE	mixed	none	none	good	good	--	good	good	good	--	--	83%	0.32	0.27
Wang & Newlin (2000)														
external LoC	mixed	none	none	good	--	--	--	good	good	--	--	4%	0.41	0.33
Study/Construct	**Delivery**	**Validity[a]**	**Reliability[b]**	**1[c]**	**2[d]**	**3[e]**	**4[f]**	**5[g]**	**6[h]**	**7[j]**	**8[k]**	**9[l]**	***r***	r_{adj}
internal LoC	mixed	none	none	good	--	--	--	good	good	--	--	4%	-0.33	-0.41
Wang & Newlin (2002)														
learning SE	mixed	none	none	good	good	--	--	--	good	--	--	--	0.32	0.17
technology SE	mixed	none	none	good	good	--	--	--	good	--	--	--	0.24	0.09
Wang et al. (2008)														
intrinsic motiv	unclear	none	attempt	good	--	--	good	good	good	--	--	--	0.16	0.22
technology SE	unclear	none	attempt	good	--	--	--	good	good	--	--	--	0.05	0.11
external LoC	unclear	none	attempt	good	--	--	--	good	good	--	--	--	-0.07	-0.01
Warren et al. (2020)														
learning SE	mixed	none	none	--	good	good	--	--	--	--	--	74%	0.46	0.34
Wipawayangkool et al. (2022)														
technology SE	unclear	good	good	good	good	--	good	good	good	good	good	--	-0.01	-0.04
Wuellner (2015)														
intrinsic motiv	asynch	none	none	--	good	--	good	--	good	--	--	18%	-0.41	-0.53
Zhang et al. (2023)														
technology SE	unclear	good	good	good	good	--	good	good	good	good	good	16%	0.13	0.13
Zhou & Wang (2019)														
mastery approach	synch	good	good	good	good	--	good	good	good	good	good	31%	0.61	0.61

Study/Construct	Delivery	Validity[a]	Reliability[b]	1[c]	2[d]	3[e]	4[f]	5[g]	6[h]	7[i]	8[j]	9[k]	*r*	r_{adj}
mastery avoidance	synch	good	good	good	good	--	good	good	good	good	good	31%	0.44	0.44
performance approach	synch	good	good	good	good	--	good	good	good	good	good	31%	0.35	0.35
performance avoidance	synch	good	good	good	good	--	good	good	good	good	good	31%	-0.05	-0.05
Zimmerman (2017)														
learning SE	mixed	none	good	good	good	--	good	good	good	good	--	34%	-0.02	-0.02
technology SE	mixed	none	good	good	good	--	good	good	good	good	--	35%	-0.10	-0.10
learning SE	mixed	none	good	good	good	--	good	good	good	good	--	36%	0.12	0.12

Note: For Construct: SE=self-efficacy; LoC=locus of control; motiv=motivation; RAI=relative autonomy index; Delivery=synchronicity --=bad, unclear, or not reported (blank); r=raw effect size; r_{adj}=Hunter's r, adjusted for sample size and quality

[a]*Measurement Reporting Quality - Validity*

[b]*Measurement Reporting Quality - Reliability*

[c]*Study Quality - Reliability: Clear presentation*

[d]*Study Quality - Reliability: Clear methods and results*

[e]*Study Quality - Reliability: Ethics procedures described*

[f]*Study Quality - Validity: Methods align with scope of research*

[g]*Study Quality - Validity: Free from bias*

[h]*Study Quality - Validity: Explicit and transparent conclusions*

[i]*Study Quality - Applicable to setting (higher education)*

[j]*Study Quality - Confident in findings*

[k]*Study Quality - Attrition for sample (percentage*