

Protocol for Meta-Analyses Examining Personalized Adaptive Learning in Undergraduate Mathematics

Abstract

This protocol is for a study that examines personalized adaptive learning (PAL) within undergraduate students with a focus on mathematics outcomes. This protocol addresses two meta-analytic studies. 1) Examination of PAL within the population of undergraduate students enrolled in mathematics courses broadly (across all mathematics courses with type of course as a moderator) and 2) specifically in college algebra, which is a STEM pipeline mathematics course. Undergraduate mathematics courses can be roadblocks to college degree attainment. PAL offers the opportunity to develop mathematical skills in a personalized way, while enabling instructors to identify students' struggle areas and take remedial actions.

Partnering with academic librarians, databases likely to index relevant studies were identified, the search strategy was refined, and keywords and subject indexing terms were identified. Titles and abstracts were independently screened by two reviewers. Full-text review is in progress. Retained studies will have data extracted that will allow the results to be meta-analyzed.

Standardized mean difference will be the main effect size computed. Multilevel meta-analysis will be used to address the hierarchical data structure. Robust variance estimation will be used to estimate standard errors and hypothesis tests, addressing dependencies of effect sizes within studies. Prediction intervals for effect sizes, for study-average effect sizes, and heterogeneity statistics will be reported. Subgroup analyses will evaluate the relationship between moderators and the magnitude of the effect size. Cochrane's criteria for assessing risk of bias in studies with a separate control group will be applied. Reporting bias will be assessed visually with funnel plots.

Word count: 250

Protocol for Meta-Analyses Examining Personalized Adaptive Learning in Undergraduate Mathematics

Keywords: Personalized adaptive learning, mathematics, postsecondary, meta-analysis, systematic review

Highlights:

- Undergraduate mathematics courses can be roadblocks to college degree attainment. Personalized adaptive learning (PAL) offers the opportunity to develop mathematical skills in a personalized way, while enabling instructors to identify students' struggle areas and take prompt remedial actions.
- PAL is increasingly adopted as an instructional tool for mathematics to assist in addressing the large proportion of college students who are underprepared for college-level mathematics and the resulting low success rate for students placed into developmental courses.
- Very few syntheses/systematic reviews/meta-analyses have examined the extent of effectiveness of PAL on academic or other outcomes. Meta-analysis uses effect sizes, which measure the magnitude of the effect of PAL on math outcomes, as the unit of observation. This study will examine PAL within a population that has yet to be fully explored with meta-analytic methods, that of undergraduate students, with a focus on mathematics outcomes. The study will also address the limitations of previous reviews related to personalized learning (e.g., examining research quality and expanding the publication period, databases, and publication type).
- The primary meta-analysis (study 1) explores undergraduate mathematics broadly, and the secondary meta-analysis (study 2) includes a subset of studies identified in study 1 to examine college algebra, a STEM pipeline mathematics course, more specifically.

Protocol for Meta-Analyses Examining Personalized Adaptive Learning in Undergraduate Mathematics

Personalized adaptive learning (PAL) is a software platform approach to instruction that individualizes the learning experience for each student based on their demonstrated knowledge and skills (Center for Distributed Learning). Through PAL, unique learner paths are created through the implementation of intelligent learning systems, integration of learner preferences, and analyses of individual learning data (Xie et al., 2019). Advocates argue that PAL provides tools which allow students to succeed (Jones & Casey, 2015), and researchers who have studied PAL have found it to be effective in increasing mastery of content and predicting final course grades (Simon-Campbell & Phelan, 2016). PAL may be particularly beneficial to postsecondary students enrolled in courses such as college algebra and calculus, as they serve as gatekeeper courses, especially for those majoring in science, technology, engineering, or mathematics (STEM) fields (Rech & Harrington, 2000; Toven-Lindsey et al., 2015).

Within the United States, personalized adaptive learning (PAL) has been increasingly adopted in the last ten years, but arguably more attention has been paid to it as an instructional strategy since the passing of the Every Student Succeeds Act (ESSA), in which schools are encouraged to increase student access to rigorous PAL. Highlighted within ESSA is the need for educational agencies (both state and local) to develop innovative learning environments personalized to the needs of the students which utilize modern technology, adopt flexible instructional practices, and are aligned to Universal Design for Learning (UDL) principles (Every Student Succeeds Act, 2015). This study is poised to provide insight into PAL as an instructional strategy in undergraduate mathematics courses, including gatekeeper courses, that may inform tools that can be used to assist STEM pathway completion in higher education.

Purpose of the Study

This protocol is for a study whose aim is to examine the extent to which PAL is effective (as measured by effect size) in relation to mathematics outcomes in undergraduate mathematics courses. This project will specifically focus on two meta-analytic studies. *Meta-analysis 1* will examine PAL in undergraduate mathematics broadly (across all courses, with content of mathematics course (e.g., liberal arts math, algebra, trigonometry, pre-calculus, calculus) as a moderator. *Meta-analysis 2* will examine a subset of those studies reporting on PAL outcomes from college algebra, a STEM pipeline mathematics course, given the struggle many students have with prerequisite math courses that are required to enter and successfully complete STEM degrees (Ganter & Haver, 2011). Both studies will examine moderation of institutional factors, course-related factors, PAL-related factors, student-related factors, and study-related factors. Each study will examine the overall student population as well as underserved students who struggle most with the math curriculum.

Very few syntheses, systematic reviews, and/or meta-analyses have examined the extent of effectiveness of PAL on academic or other outcomes. Those that have (Xie et al., 2019; Zhang et al., 2020) used restrictive search criteria (i.e., databases searched, publication type, publication period), failed to examine research quality, and failed to present meta-analytic results. Also, the study that was broader in search (but still restrictive) (i.e., Zhang et al., 2020), was restricted to

pre-kindergarten through high school. This study will address limitations observed in previous reviews related to personalized learning (e.g., examining research quality and expanding the publication period, databases, and publication type).

The studies will focus on mathematics courses (e.g., college algebra, calculus) since they are often reported to be a major stumbling block to both entering and completing a STEM degree (Kersaint et al., 2011; President's Council of Advisors on Science and Technology, 2012). Additionally, mathematics courses traditionally have a high failure and withdrawal (i.e., DFW) rates for most non-STEM disciplines as well (Hagedorn & DuBray, 2010). DFW rate is the percent of students who receive a D or F or withdraw from a course, which negatively impacts college progression and possibly financial aid requirements. Thus, undergraduate mathematics courses can serve as roadblocks to college degree attainment for many students, regardless of degree, and may particularly impact underrepresented students (Chen, 2013; National Academies of Sciences, 2016). Additionally, the Mathematical Association of America's committee on Curriculum Renewal Across the First Two Years (CRAFTY) derived a set of recommendations to remodel the college algebra course (CRAFTY, 2007). Since then, universities and colleges have invested considerable efforts towards addressing the key challenges in teaching college mathematics (Gordon & Gordon, 2021). In this regard, the use of PAL offers students the opportunity to develop mathematical skills in a personalized way while enabling instructors to identify students' struggle areas early and take appropriate remedial action. Identifying students' deficits early in the semester provides the best case for successful course completion. Thus, as an educational tool, PAL aligns with CRAFTY's guidelines.

Research Questions

The primary meta-analysis (study 1) explores undergraduate mathematics broadly, and the secondary meta-analysis (study 2) includes a subset of studies identified in study 1 to examine college algebra, a STEM pipeline mathematics course, more specifically. The primary research question is: What is the average effect size of personalized adaptive learning on *undergraduate mathematics outcomes broadly (study 1)* and specifically on *college algebra (study 2)*, based on the empirical literature? The secondary research question is: To what extent is the effect size of personalized adaptive learning on mathematics outcomes moderated by *institutional-related factors* (e.g., institution type), *course-related factors* (e.g., course modality, general education course [*general education for study 1 only*]), *PAL-related factors* (e.g., PAL 'dosage'—proportion of PAL course components), *student-related factors* (e.g., gender, low income or Pell-eligible, first generation, ethnicity), and *study-related factors* (e.g., research quality, publication bias)?

Significance of the Study

While PAL is increasingly adopted in postsecondary courses, there are very few systematic reviews or syntheses that examine the impact of PAL and none that have been found that examine PAL specifically for mathematics. This proposed research will demonstrate the extent to which personalized adaptive learning (PAL) may be effective toward increasing student outcomes in undergraduate mathematics courses. This research will clarify where (e.g., types of institutions) and how PAL is implemented in postsecondary mathematics courses and will evaluate the extent to which PAL may relate to positive cognitive and affective mathematics

student outcomes. Should results suggest PAL is effective, recommendations on best methods of using PAL for mathematics instruction will be proposed. These results will be beneficial to postsecondary mathematics instructional faculty, instructional designers, and administrators in providing valuable information for future design and implementation of postsecondary mathematics courses. Results will further help educators to understand factors that may moderate the success of PAL.

Successful STEM pathway completion has been a persistent challenge in higher education (Toven-Lindsey et al., 2015). Less than 40% of STEM majors complete a degree within six years (President's Council of Advisors on Science and Technology, 2012) with 48% of bachelor's students and 69% of associate's STEM entrants from 2003 to 2009 leaving the field by spring 2009 (Chen & Soldner, 2013). Beginning with low completion in developmental math (Bryk & Treisman, 2010) (50% at 2-year college; 58% at 4-year institutions) (Rutschow et al., 2019) and low success (A, B, C grade) in college algebra (less than 50%) (Ganter & Haver, 2011), students struggle with the prerequisite math required to enter and successfully complete STEM degrees. Research suggests that student attrition in the sciences occurs within two years of taking these gatekeeper courses (Chang et al., 2008). In addition, poor college performance and high withdrawal and failure in STEM courses are associated with an increased probability of dropping out of college for bachelor's and associate's STEM entrants (Chen & Soldner, 2013).

Further, while underrepresented minorities are as likely to enter a STEM degree program as non-minority peers (Crisp et al., 2009), they are less likely to attain a STEM degree (Hurtado et al., 2010). Women and minorities still represent less than 25% of computer science, physical science, and engineering graduates and less than 50% of graduates in mathematics and statistics (National Center for Science and Engineering Statistics, 2021). Experts call for a reassessment of curricula, pathways, and barriers, and a focus on evidenced-based instructional techniques to help address the struggle of undergraduates in completing STEM disciplines (Saxe & Braddy, 2015). This project will help address this critical need by methodologically examining research on PAL, specific to mathematics curriculum, and identifying where, when, and for whom PAL is helpful in relation to mathematics outcomes.

The Intervention

Personalized adaptive learning (PAL) has emerged as an instructional method due to the increase in technology for big data (Hey et al., 2009). PAL implements intelligent learning systems, integrates learner preferences, and analyzes individual learning data to create a unique path for each learner (Xie et al., 2019). Core elements of PAL are individual characteristics and performance, personal development, and adaptive adjustment (Peng et al., 2019). The learner is fundamental to the system and thus the course design (Aroyo et al., 2006). User preferences (e.g., language) and current knowledge state of the learner may be recorded in the system as a model of each student and this information overlays the domain model to record the current status of the learner in terms of their progress and knowledge of domain concepts. These "application dependent user models," that include a learner's knowledge state and preferences, are integrated with the profile that comprises the available information known about the learner. This becomes the basis for adapting how the content is presented to the learner (Aroyo et al., 2006, p. 6). Recent research indicates enhanced learning with PAL instruction adapts to students'

prior knowledge, learning strategies and challenges, affect and motivation, and ability to regulate their own learning (Aleven et al., 2017). It can also provide the basis for learning analytics dashboards that can inform instruction, identify modules where students struggle, and alert instructors to struggling students so that remedial action can be taken.

The U.S. National Education Technology Plan defines personalized adaptive learning as “instruction in which the pace of learning and the instructional approach are optimized for the needs of each learner. Learning objectives, instructional approaches, and instructional content (and its sequencing) may all vary based on learner needs. In addition, learning activities are meaningful and relevant to learners, driven by their interests, and often self-initiated” (2017, p. 9). Although some scholars differentiate ‘adaptive learning’ from ‘personalized learning’ based on the approach for catering to the learners’ needs (Xie et al., 2019), many studies use the terms interchangeably (e.g., Aroyo et al., 2006; Gobel et al., 2010; Lin et al., 2013). PAL is used in conjunction with various course modalities (e.g., online, blended, face-to-face) and instructional models (e.g., supplemental instruction, asynchronous learning, synchronous learning, emporium model). This study will include the broad use of PAL to capture, identify and detail any models which succeed in improving student success in STEM.

PAL within Postsecondary Institutions

Although research suggests students’ perceptions of PAL are positive (Howlin, 2014; Johnson & Samora, 2016; Simon-Campbell & Phelan, 2016), research on cognitive outcomes is limited regarding its effectiveness. Based on an early evaluation of Gates Foundation PAL grantees, the odds of course completion did not change based on PAL, although four of 15 projects resulted in higher average course grades. Further, the seven available side-by-side comparisons of learning assessments showed modest but positive average impact for PAL (Yarnall et al., 2016). In providing remedial instruction in biology, chemistry, information literacy, and mathematics to first-year pharmacy students, Liu et al. (2017) found that adaptive learning was successful only for chemistry. However, there is some evidence that the effectiveness of PAL may be moderated by course, institutional type, and student subgroups. Moving from lecture to blended instruction with PAL has been found to positively impact student learning (Yarnall et al., 2016). Within online classes, switching to PAL from online without PAL has resulted in small positive impact on course grades (Yarnall et al., 2016). Comparing student outcomes using PAL in statistics courses at four-year and two-year institutions, course grades and passing rates, as well as statistical competency, statistically increased for students at four-year institutions only but there was no difference based on student subgroups (e.g., first-generation, Pell grant) (Joo & Spies, 2019).

Within content areas, PAL is increasingly adopted as an instructional tool for mathematics, and specifically developmental mathematics (Weiss & Headlam, 2018), to assist in addressing the large proportion of college students who are underprepared for college-level mathematics (Chen, 2016) and the resulting low success rate for students placed into developmental courses (Bailey et al., 2016; Wheeler & Bray, 2017). In a randomized trial, researchers found that PAL students (relative to control) were more likely to earn one or more developmental math credit and complete a higher proportion of the math sequence (Weiss & Headlam, 2018). In follow-up semesters, PAL students were more likely to persist to the second semester, enroll in a

mathematics class, and earn at least one math credit. However, there were similar success rates between PAL and control students who completed the first one-half of the developmental math sequence, who were deemed college-ready through completing the entire developmental mathematics sequence, and who passed their first college-level course in math (Weiss & Headlam, 2018).

Faculty increasingly utilize digital tools and courseware, with adoption higher in those who teach introductory courses. This trend started prior to the pandemic but has accelerated with more faculty teaching online. Fox et al. (2021) found that 86% of faculty teaching introductory courses used at least one digital tool (e-text, instructional tools, or courseware such as adaptive learning). Most recently, the Every Learner Everywhere (ELE) network supported the implementation of adaptive learning in two- and four-year colleges in high enrollment “gateway” courses. First year reports from participating schools indicated that 89% of faculty using adaptive learning reported gains in student learning. The ability to craft practice opportunities that adapt and are personalized to student ability and engagement, as well as quick, actionable feedback for students and instructors, were positives. The potential to reduce equity gaps, if students have appropriate technology and support, was also a strength that faculty mentioned. The American Public Land Universities (APLU) Personalized Learning Consortium also supported implementation of adaptive learning. Case studies from participating lighthouse institutions utilizing adaptive courseware for mathematics found increases in student success rates and positive outcomes in addressing the wide range of needs and varying knowledge of students in gateway and introductory math courses (Achieving the Dream, 2021a, 2021b, 2021c, 2021d, 2021e, 2021f). These studies also indicate the need to examine results by student and course demographics to determine disparity, which we do in this study.

Reviews, Syntheses, and Meta-Analyses on PAL

Very few syntheses, systematic reviews, and meta-analyses have examined PAL. Yarnall et al. (2016) conducted a meta-analysis of the Bill & Melinda Gates Foundation funded Adaptive Learning Market Acceleration Program (ALMAP) grantees which included data collected from over 19,500 students and 280 instructors. This nationally cited project focused on learning, cost, and satisfaction outcomes, finding slightly higher course grades in only four of 15 implementations and a positive impact on course completion odds in only two of 16 side-by-side case comparisons. The average impact of PAL for seven controlled side-by-side comparisons of learning assessment scores was modest, but significantly positive. They used meta-analysis techniques computing effect sizes on student outcomes and odds ratios for binary measures, transforming the latter into effect size metrics for comparisons. Their study was limited to only grant recipients, and further restricted by limited comparable data across the participants which had few STEM participants.

Xie et al. (2019), conducted a systematic review of journal publications focused on adaptive/personalized learning without restriction to age or type of learner (including studies from elementary school to college to working adult) or outcome (including studies with outcomes related to affect, cognition, skills, behavior, and review/conceptual articles). Their search did not broadly consider search phrases for PAL (e.g., only two search terms, “adaptive learning” or “personalized learning;” Web of Science database only; publication period, 2007-

2017; publication type=article; and category=education/educational research), resulting in only 161 articles in the query, of which 70 met the inclusion criteria (32 of which included postsecondary students). Additionally, the authors did not conduct a meta-analysis of study effect sizes, and the authors did not code for research quality, which will be addressed in the proposed study.

Zhang et al. (2020) conducted a research synthesis of empirical studies of personalized learning implementation in pre-kindergarten to grade 12 settings, without restriction to outcome (including studies with academic, engagement, metacognitive, and attitude toward learning). Their search was also restrictive (e.g., publication period, 2006-2019; databases limited to ERIC, OmniFile Full Text Select, and Academic Search Complete, and Web of Science; publication type = peer reviewed journal), resulting in 2,131 identified studies and 71 articles which ultimately met inclusion criteria. The authors did not conduct a meta-analysis of study effect sizes, and the authors did not code for research quality, which will be addressed in the proposed study.

Methods

Criteria for Study Inclusion and Exclusion

Inclusion and exclusion criteria for studies are detailed in Table 1 and encompassed type of citation (i.e., study), language (English), country where the study was completed (United States), level of coursework (undergraduate), educational setting (course), type of course (mathematics), study design (randomized or quasi-experimental), intervention (PAL), and outcome (cognitive or affective mathematics). Articles from the initial search were included if they met the criteria identified.

Table 1
Criteria for Study Inclusion and Exclusion

Category	Criteria
Type of citation	The citation must denote a study. Newspaper articles, blogs, tweets, slide decks, and similar outlets that reference studies as secondary (not primary) sources were not included.
Language	Studies must be reported in the English language. Given the resource and time limitation for the project, inclusion of non-English studies was not possible. However, restriction by language was not imposed at the search stage but rather at the study selection stage (i.e., abstract and title screening and full text eligibility screening) with English language as an eligibility criterion (Higgins et al., 2022).
Country	The study must take place in the U.S., U.S. territories (American Samoa, Guam, Northern Mariana Islands, Puerto Rico, Virgin Islands), freely associated states (Federated States of Micronesia, Republic of Palau, Republic of the Marshall Islands) or Washington DC. Studies conducted in any other country are excluded given that the postsecondary education systems may not be comparable to those in the U.S.

Postsecondary level	Only studies that include undergraduate courses as the sample will be included. This excludes PK-12 and graduate courses, which are not the focus of the study. Undergraduate courses are the focus of this study as mathematics courses at the undergraduate level are gatekeeper courses for advancing in STEM related fields (Chen, 2013; National Academies of Sciences, 2016).
Educational setting	The study must take place in an educational setting, including a face-to face, blended, or online course. Studies that are laboratory-type experiments in settings that are not part of a course are excluded. Non-credit courses, MOOCs, Coursera, and similar that are not part of undergraduate curriculum are excluded. Bootcamp and summer preparatory programs that are non-credit bearing are excluded. Credit-bearing courses are the focus of the study as they directly influence student success in their undergraduate programs.
Type of course	The study must occur in a mathematics course (e.g., liberal arts math, algebra, business statistics, trigonometry, pre-calculus, calculus). Studies that occur in any other type of course [e.g., mathematics education (teacher education courses), science, engineering] are not eligible. Mathematics courses at the undergraduate level are gatekeeper courses for advancing in STEM related fields (e.g., Chen, 2013; National Academies of Sciences, 2016), and thus examining interventions such as PAL that are used within undergraduate math courses is important.
Study design	The study design must be a group design randomized controlled trial or a quasi-experimental design. Qualitative, single subject, and single group research designs will be excluded. Studies that do not include a comparison group or condition, or studies involving meta-analysis or literature review are not eligible. Studies such as pre-post designs that do not include a comparison group or condition are excluded as they may produce biased results given the independence between measurement occasions (Cuijpers et al., 2017). Meta-analyses are excluded as they are secondary sources. Literature reviews are excluded as they do not contain study results details.
Intervention	Only empirical investigations with observed data associated with a personalized adaptive learning activity will be included. The inclusion criteria for PAL will consider online PAL systems (e.g., ALEKS, McGraw-Hill Connect/LearnSmart, IMathAS, Knewton Alta, MyLab, Lumen Waymaker, IMathAS, Inquizitive, MyMathLab, Wiley Plus, Acrobatiq, Smart Sparrow, CogBooks, and similar) as well as ‘build your own’ PAL components/assignments which may be faculty-created using platforms such as Realizeit or Acrobatiq. The ‘build your own’ systems are still part of online PAL systems but require faculty to build the online modules (i.e., create the content and/or assessments) as opposed to the publisher providing modules. Research on instructional methods that do not provide any adaptive content to adjust for individual student learning to optimize student learning will be excluded to ensure the focus of the intervention is PAL. Articles focused on faculty development on how to use PAL, theoretical studies, policy briefs, and studies with simulated data (e.g., Monte Carlo simulations), as examples,

	will be excluded. This study also excludes studies that explain only the development of PAL platforms or PAL courses, but which do not include requisite empirical research.
Outcome	The study must incorporate a cognitive or affective mathematics student measure. This math-related measure may be students' mathematics knowledge or academic skills or other mathematics-related topic such as engagement in learning mathematics, anxiety towards math, or attitude toward learning mathematics. Measures on teachers, faculty, staff, and individuals other than students enrolled in the course are excluded and measures unrelated to mathematics are excluded. Those excluded measures will help ensure that the study's focus is related only to mathematics.

Search Strategy

The Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) 2020 was followed to describe the search strategy, providing information on identification of terms used in the search (e.g., keywords, subject index terms) as well as the peer review process to assess the search strategy (Moher et al., 2010; Page et al., 2021; Rethlefsen et al., 2021). As noted as a best practice (Higgins et al., 2022), development and operationalization of the search strategy began with consideration of terms to identify the population, intervention, comparison, and outcomes (PICO) (Richardson et al., 1995), which was framed from the research question, as well as terms to identify the type of study design (i.e., randomized trial or quasi-experimental) (Higgins et al., 2022) which is the PICOS framework (Centre for Reviews and Dissemination, 2009). Drawing on their expertise in various aspects related to the project, the research team developed the initial list of keywords. Additional keywords were considered based on terms used in previous PAL reviews (Xie et al., 2019; Zhang et al., 2020). Once this list was developed, members of the project's advisory board (who had expertise in personalized learning, mathematics, and systematic review methodology) assisted in identifying additional keywords and modifiers to maximize the depth and breadth of research terms. Additionally, we searched publisher websites of those with known personalized learning platforms/curricula to identify additional search terms based on how the publishers were referring to PAL.

Partnering with an academic librarian knowledgeable in systematic reviews, we identified databases that were likely to index relevant studies and further refined the search strategies by determining additional keywords and subject indexing terms (e.g. via database thesaurus). This search strategy was peer-reviewed by another librarian, prior to executing the search, who offered additional feedback to maximize the potential of the search in identifying relevant studies, a best practice as noted by the PRISMA-S Extension (Rethlefsen et al., 2021). The Peer Review of Electronic Search Strategies (PRESS) Evidence-Based Checklist (McGowan et al., 2016b), which assesses technical accuracy, the extent the search strategy addresses relevant elements of the protocol, and appropriate interpretation of the research question, was used to evaluate the important elements in the search strategy peer review process (McGowan et al., 2016a, 2016b). The use of a checklist such as PRESS has been found to improve the quality of systematic review literature searches (Relevo & Paynter, 2012; Spry et al., 2013) and assess the adequacy of the search strategy (Higgins et al., 2022). A review of the search strategy on the dimensions of the PRESS suggests no missing dimensions that are not adequately explained and

provides evidence that the search is complete (Higgins et al., 2022).

Identified keywords, using the Patient, Intervention, Comparison, Outcome, and Study Design (PICOS) framework (Centre for Reviews and Dissemination, 2009), are listed in Table 2. Concepts within topic area were combined by the OR Boolean operator, and concepts across topic areas were combined with the AND Boolean operator. Truncation and wildcards were used as needed to open the range of terms identified (e.g., universit* to allow variants including university and universities to be captured). Searches were customized to each database. It is recognized that college* may identify studies that include “college student*” and thus “college student*” may be repetitive. The purpose for including both terms is that in some databases “college” was used as a subject descriptor/subject term. In most searches, “college*” was also used as a truncated keyword and “college student*” as a truncated keyword phrase. All databases were independently searched by two research team members.

Table 2
Keywords by PICOS Topic Area

P	college* OR “college student*” OR postsecondary OR undergraduate* OR “undergraduate student*” OR universit*
I	“adaptive assess*” OR “adaptive course*” OR “adaptive distance education” OR “adaptive distance learn*” OR “adaptive elearn*” OR “adaptive e-learn*” OR “adaptive instruct*” OR “adaptive learn*” OR “adaptive practice” OR “adaptive test*” OR “individualized elearn*” OR “individualized e-learn*” OR “individualized instruction” OR “individualized learn*” OR “intelligent tutor* system*” OR “online adaptive learn*” OR “personalised adaptive learn*” OR “personalised distance education” OR “personalised distance learn*” OR “personalised e-learn*” OR “personalised elearn*” OR “personalised instruction” OR “personalised learn*” OR “personalised online learn*” OR “personalized adaptive learn*” OR “personalized distance education” OR “personalized distance learn*” OR “personalized e-learn*” OR “personalized elearn*” OR “personalized instruction” OR “personalized learn*” OR “personalized online learn*” OR Acrobatiq OR ALEKS OR CogBooks OR IMathAS OR Inquizitive OR “Knewton Alta” OR LearnSmart OR “Lumen Waymaker” OR “McGraw Hill Connect*” OR MyLab OR MyMathLab OR RealizeIt OR “Smart learn*” OR “Smart Sparrow” OR “Wiley Plus”
C	NOT APPLICABLE
O	algebra OR calculus OR “differential equations” OR math* OR “math* education” OR precalculus OR trigonometry
S	“clinical trial” OR compar* OR contrast OR “control group*” OR correlat* OR “effect size*” OR experim* OR intervention* OR “matched group*” OR program* OR QED OR “quasi experim*” OR quasiexperim* OR quasi-experim* OR “random* control*” OR “random* design” OR “random clinical” OR “random* trial*” OR RCT OR strateg* OR therap* OR treatment*

Recognizing that each database has a tailored search strategy, an example of a representative search statement appears similar to the following (for brevity, the customized search strings by database are not presented):

(college* OR “college student*” OR postsecondary OR undergraduate* OR “undergraduate student*” OR universit*)

AND

(algebra OR calculus OR “differential equations” OR math* OR “math* education” OR precalculus OR trigonometry)

AND

(“clinical trial” OR compar* OR contrast OR “control group*” OR correlat* OR “effect size*” OR experim* OR intervention* OR “matched group*” OR program* OR QED OR “quasi experim*” OR quasiexperim* OR quasi-experim* OR “random* control*” OR “random* design” OR “random clinical” OR “random* trial*” OR RCT OR strateg* OR therap* OR treatment*)

AND

(“adaptive assess*” OR “adaptive course*” OR “adaptive distance education” OR “adaptive distance learn*” OR “adaptive elearn*” OR “adaptive e-learn*” OR “adaptive instruct*” OR “adaptive learn*” OR “adaptive practice” OR “adaptive test*” OR “individualized elearn*” OR “individualized e-learn*” OR “individualized instruction” OR “individualized learn*” OR “intelligent tutor* system*” OR “online adaptive learn*” OR “personalised adaptive learn*” OR “personalised distance education” OR “personalised distance learn*” OR “personalised e-learn*” OR “personalised elearn*” OR “personalised instruction” OR “personalised learn*” OR “personalised online learn*” OR “personalized adaptive learn*” OR “personalized distance education” OR “personalized distance learn*” OR “personalized e-learn*” OR “personalized elearn*” OR “personalized instruction” OR “personalized learn*” OR “personalized online learn*” OR Acrobatiq OR ALEKS OR CogBooks OR IMathAS OR Inquizitive OR “Knewton Alta” OR LearnSmart OR “Lumen Waymaker” OR “McGraw Hill Connect*” OR MyLab OR MyMathLab OR RealizeIt OR “Smart learn*” OR “Smart Sparrow” OR “Wiley Plus”)

Electronic Databases

Best practice in systematic reviews recommends conducting a search that is as extensive as possible, with the purpose of reducing the risk of reporting bias and increasing the probability of identifying relevant studies (Higgins et al., 2022). Thus, the literature search was multi-pronged to ensure depth and breadth of coverage to include: a) keyword searches of applicable databases and grey literature search (Higgins et al., 2022) including b) conference proceedings/program search; c) citation mining (i.e., searching reference lists); and d) contact with authors.

The databases included those most likely to index PAL studies. Given the cross-disciplinary nature of the topic, databases from multiple discipline areas were included (Frandsen et al., 2019; Stevinson & Lawlor, 2004; Wallace et al., 1997) that index related topics (e.g., education, mathematics, e-learning). The specific databases searched included the following:

1. Academic Search Premier (via EBSCO Host)
2. ACM Digital Library
3. APA PsycInfo (via EBSCO Host)
4. Compendex (Engineering Village)
5. Education Source (via EBSCO Host)
6. ERIC (via EBSCO Host)
7. IEEE Xplore
8. Inspec (Engineering Village)
9. ProQuest Dissertations & Theses Global (via ProQuest)
10. ProQuest SciTech Premium Collection (via ProQuest)
11. ProQuest Social Science Premium Collection (via ProQuest)
12. Science Direct (suggest do initial search and limit) (via Elsevier)
13. Web of Science (Clarivate) (all WOS including Conference Proceedings Citation Index)

Zotero was the citation management software used to organize and manipulate citations. A search, using the terms in Table 2 and specialized to each database using thesaurus terms was used to search each of the databases without date, language, or publication/format/document type restriction. Not restricting searches by publication type and language reduces the possibility of publication and/or language bias and not restricting by document type allows for documents to be included that may contain information not cited elsewhere (Higgins et al., 2022). As noted in the inclusion criteria, given the resource and time limitation for the project, inclusion of non-English studies was not possible. However, restriction by language was not imposed at the search stage but rather at the study selection stage with English language as an eligibility criterion (Higgins et al., 2022). The search, which has commenced, was tailored to each database, utilizing the vocabulary and thesaurus in each database.

Other Resources

Using identified keywords (Table 2), conference programs/repositories for applicable conferences not indexed in the previous databases will be searched including: AMATYC abstracts, American Educational Research Association, EDUCAUSE, ICTCM abstracts, IEEE Frontiers in Education, Joint Meeting abstracts, Learning Analytics and Knowledge (LAK), MathFest abstracts, Online Learning Consortium, PME proceedings, PME-NA proceedings.

To extend the search further, references cited in eligible studies will be reviewed including backward reference searching (i.e., identifying works cited in the eligible studies) and forward reference searching (i.e., identifying articles that cite eligible studies). Searching reference lists as an additional source for relevant studies is noted by Cochrane as mandatory to decrease the possibility of publication bias and assist in identifying additional relevant studies (Higgins et al., 2022). Citation mining that results in few or no additional studies also provides evidence that the search strategy performed adequately (Higgins et al., 2022). Databases that provide citation mining include PsycInfo and Web of Science. Google Scholar has a built-in citation mining tool and will also be used to access eligible ‘cited by’ studies.

Author Query

Authors who have published eligible studies within the past five years (to concentrate on scholars more likely to have a current research focus in this area) will be contacted. We will invite these authors to share related studies, published and unpublished, and recommend additional authors who may have applicable unpublished studies.

Data Collection and Analysis

Citation screening is being performed in a multi-stage process. Stage 1: Titles and abstracts were screened by reviewers. Articles excluded at this stage were those titles and abstracts that clearly violated one or more inclusion criteria (previously noted). Stage 2: The second stage of screening is full-text review. At this stage, the entire text of articles is reviewed. Only articles that meet all inclusion criteria are retained. Retained studies will have data extracted which will allow the results to be meta-analyzed. Stage 3 (as needed): At stage 3, authors of studies that did not have sufficient data available in the text will be contacted with a request to supply the requisite data. Should author queries fail to provide the requisite data, those studies will be excluded and noted accordingly to delineate them from studies that were excluded based on failing to meet all the inclusionary criteria.

Abstract and Title Screening

A best practice in abstract and title screening is to use a text-mining abstract screening application that provides for easy project management (Polanin et al., 2019). For this project, Covidence is being used. Two researchers independently screened titles and abstracts, which is highlighted as a best practice in meta-analysis (Waffenschmidt et al., 2019). One of the two researchers is one of the project team leads. Discrepancies are resolved via discussion. Abstract and title screening training was conducted prior to commencing with screening (see Training, Data Extraction, and Management).

Full Text Eligibility Screening

After abstract and title screening, full texts of eligible studies identified in the screening stage will be located, uploaded to Covidence, and reviewed. Applying the study inclusion criteria, two reviewers will independently screen the full text to determine inclusion or exclusion. As recommended by Cochrane (Li et al., 2023), the use of more than one reviewer reduces the likelihood of errors and introduction of biases. For studies identified as eligible at the full text screening stage, quantitative data needed for computing effect sizes such as means and standard deviations, correlations, test statistic values, and sample sizes, will be extracted.

Training, Data Extraction, and Management

The lead researchers collaboratively developed a screening manual which includes a codebook for all characteristics to be studied in the abstract screening stage. In training screeners, case studies of a small sample ($n = 10$) of abstracts from eligible and ineligible studies were reviewed together. After completing training, all screeners independently screened abstracts for 25 studies. The lead researchers reviewed point-by-point results, noted areas of disagreement, and met with the screeners to discuss and remediate. At this point, any revisions to the screening manual to assist in clarification were made. This cycle commenced again with another set of 25 abstracts for all screeners to independently screen; the lead researchers reviewed, discussed, and

remediated; and any final revisions to the manual were made. Remaining eligible studies were screened by two coders, one of which was one of the two team leads. Content coding less than exact agreement will be discussed to reach consensus. Once screening commenced, refinements to clarify the criteria was made, as needed, and the screening team was appropriately trained.

A similar training process is being followed for full-text screening training. Covidence is used to manage citations for full-text eligibility screening, and MetaReviewer will be used for data extraction. Coders will review the screening manual and data extraction tool. Three full-text studies will be independently reviewed by all coders. Disagreements will be discussed by the coding team. Three more full-text studies will be coded independently, with point-by-point discussion by the team again. This cycle of review will continue until 80% interrater reliability is reached. Remaining studies will be screened by two coders, one of which will be one of the team leads. When there is less than exact agreement, discrepancies will be discussed to reach consensus. A running agenda item on the weekly team meetings will be discussion of interrater reliability and coding challenges.

Moderators that will be coded include attributes related to the sample, PAL, class, and institution as well as instrumentation properties (see Table 2). Other moderators will be included should full-text coding provide insight into other available attributes that have merit.

Table 2
Moderators

<i>Category</i>	<i>Moderator</i>	<i>Response Scale</i>
<i>Sample attributes</i>	Sex	% females and males in the class
	Race	% White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, and Other
	Hispanic	% Spanish/Hispanic/Latino
	Class rank	% freshmen, sophomore, junior, senior, post-bac, non-degree
	SES	% Pell grant recipients
	First generation	% first generation students
	Transfer	% transfer students
<i>PAL attributes</i>	Dosage	% of the course content implemented using PAL
	Weight	% of the course grade attributed to PAL assignments
	Type	Pre-fab (e.g., ALEKS), instructor-created, combo/other
<i>Course attributes</i>	Modality	Face-to-face, hybrid/blended, online, other

<i>Category</i>	<i>Moderator</i>	<i>Response Scale</i>
	General education course (<i>study 1 only</i>)	Yes, no
	Course level	Freshmen, sophomore, junior, senior
	Course (<i>study 1 only</i>)	Liberal arts, algebra (beginning, intermediate, college), trigonometry, pre-calculus, calculus
<i>Institution attributes</i>	Carnegie classification	Doctoral university, master's college or university, baccalaureate colleges, baccalaureate/associate colleges, associate's colleges, special focus institution
	Type	Public, private
	Historically Black College or University	Yes, no
	Hispanic Serving Institution (HSI)	Yes, no
	Undergraduate enrollment	Number and % of total undergraduate enrollment
	Census region	Northeast (New England, middle Atlantic), Midwest (East North Central, West North Central), South (South Atlantic, East South Center, West South Central), West (Mountain, Pacific)
<i>Instrumentation</i>	Outcome type	Cognitive (e.g., course grade, test score), affective (e.g., interest, motivation, anxiety, attitude)
	Reliability	Scale reliability evidence (e.g., Cronbach's alpha)
	Validity	Scale validity evidence
	Standardized	Yes, no

Additional data extracted from studies during full-text coding will include:

- Research questions/purpose of the study
- Research design (RCT, quasi-experimental), number assigned to condition, and unit of allocation (person or cluster)
- Inclusion/exclusion criteria
- Random assignment allocation method and blinding (e.g., participants, outcome assessors)
- Treatment/intervention characteristics (e.g., length of treatment, 'dosage')
- Sample size at all measurement occasions (e.g., baseline, post, follow-up)
- Outcomes (e.g., name, operational definitions, unit of measure, reliability/validity evidence)

Assessing Risk of Bias

Cochrane's nine standard criteria for assessing risk of bias in studies with a separate control

group will be applied (Cochrane Collaboration, 2017). These apply to randomized and non-randomized designs if there is a separate control or comparison group. Each criterion is evaluated on a three-point scale (low risk = no bias evidence; high risk = evidence of bias; unclear = not specified/insufficient data to make adjustment).

The evaluation criteria for risk of bias relate to the following (Cochrane Collaboration, 2017): 1) *random sequence generation* (e.g., low risk if a random method is described); 2) *allocation concealment* (e.g., low risk if allocation of the unit was performed by the research team and done on at the beginning of the study for all units); 3) *similar baseline outcome measures* (e.g., low risk if all units were measured prior to treatment and baseline equivalency was determined or, in the case of imbalance, adjusted analyses were performed); 4) *similar baseline characteristics* (e.g., low risk if units in all conditions are similar); 5) *incomplete outcome data* (e.g., low risk if missing would likely not bias the results); 6) *blind allocation* (e.g., low risk if outcomes assessed blindly or outcomes were objective measures); 7) *contamination protection* (e.g., low risk if fidelity to treatment); 8) *selective outcome reporting* (e.g., low risk if outcomes were not reported selectively); 9) *other bias risk* (e.g., low risk if no evidence of risk of bias from other sources).

Assessing Reporting Bias

Reporting bias arises when the nature of the results impacts dissemination (Chan et al., 2014). Publication bias, or selective publication, occurs when only studies with interesting findings are published. Non-reporting bias occurs when eligible studies included in a review are influenced by the results of hypothesis tests or the magnitude or direction of results (e.g., not reporting results or only partially reporting results such as ' $p > 0.05$ '). Our comprehensive search, which includes grey literature, contacting study authors, and searching multiple bibliographic databases among other strategies, minimizes risk of publication bias but does not completely remove it. As recommended by the Cochrane Collaboration, assessing selective under-reporting or non-reporting is the most beneficial, and seeking information from study authors may resolve under- or non-reporting concerns.

Reporting bias will be assessed visually with funnel plots which assist in identifying small-study effects. Funnel plots will be created following Sterne and Egger (2001) and recommendations by Cochrane (Page et al., 2020). Funnel plots are scatterplots, displaying precision of the study (e.g., effect size standard errors) from the studies included in the meta-analysis on the Y axis to the estimated effect size on the X axis (Sterne & Egger, 2001). Each dot in the funnel plot represents one study in the meta-analysis. Funnel plots that resemble a pyramid with points scattered randomly within the pyramid suggest unbiased studies as well as homogeneity between studies (Sterne & Egger, 2001). Effect size estimates from small studies may reflect less precision, and thus may have greater scatter at the bottom of the funnel plot; similarly, effect estimates from larger studies may reflect greater precision and reflect tighter scatter near the top of the pyramid. Authors who find funnel plot asymmetry must take care in interpreting the source of asymmetry as visual interpretation is subjective (Lin & Chu, 2018). Recommendations for interpreting results from tests that examine asymmetrical funnel plots include, among others: a) apply tests for asymmetry only when there are at least 10 studies and interpret test results in conjunction with visual interpretation of the funnel plot; b) understand that there are multiple explanations for funnel plot asymmetry, of which publication bias is only one; and c) do not compute tests for

funnel plot asymmetry when studies are similar in size (Sterne & Egger, 2001).

Effect Size Computation

Standardized mean difference (d) will be the main effect size computed. This effect size can be computed from means and standard deviations, results of t -tests and one-way ANOVA, and more, and can also be estimated from dichotomous outcomes and chi-square statistics, phi coefficients, and other statistics. Hedges correction will be used to convert d into Hedges' bias-corrected standardized mean difference g , as d may overestimate effect in small samples (Hedges, 1981). For studies reporting only correlations, Fisher's Z transformation and variance will be computed. Campbell Collaboration's meta-analysis effect size calculator will facilitate effect size computation.

Authors of studies without sufficient data available in the text to extract effect size will be contacted with a request to supply the requisite data. Should author queries fail to provide the requisite data, those studies will be excluded and noted accordingly to delineate them from studies that were excluded based on failing to meet inclusionary criteria.

Measures of Treatment Effect

Our modeling technique is similar to ordinary least squares regression in that a dependent variable is predicted by one or more independent variables. The difference is that the dependent variable is the effect size, and we also expect a hierarchical structure. Specifically, multilevel meta-analysis, aka 'mixed-effect meta-regression', will be the data analytic approach to measure treatment effect due to the expected hierarchical nature of the data. It is anticipated that many studies included in the meta-analysis will report more than one effect size based on a common pool of participants or may report effect sizes from different samples with which a common or similar research protocol was applied (Tipton & Pustejovsky, 2015), thus creating a hierarchical structure. With multilevel or mixed-effects models, we can analyze studies within a publication, allowing the examination of variation of studies within publication as well as between publications (i.e., heterogeneity between studies). A random-effects model will allow us to estimate the overall effect across all studies (i.e., intercept randomly varies across all studies), and the mixed-effects model will allow us to also examine variation in slopes between studies with the inclusion of moderators as fixed effects (Borenstein et al., 2010).

Robust variance estimation (RVE) will be used to estimate standard errors and hypothesis tests, addressing dependencies of effect sizes within studies. With RVE, standard errors are estimated using the variation in effect sizes (rather than estimating variance from the model). With RVE, multiple correlated effect sizes can be included even when the exact structure of the dependence is not known, and this guards against misspecification (Tipton & Pustejovsky, 2015). Hypotheses tests conducted with RVE hold even if effect sizes are non-normally distributed, variances are incorrect, and weights are not inverse variance (however, heterogeneity statistics will only be approximate should the assumptions not hold).

Prediction intervals for effect sizes and for study-average effect sizes, as well as heterogeneity statistics (i.e., I^2 , proportion of observed variation that is 'true' variation), will also be reported. Prediction intervals provide a range of true effect sizes (or study-average effect sizes) found in

the population (i.e., a measure of heterogeneity). R packages, *metafor* and *clubSandwich*, will be used to estimate the models and RVE, respectively.

Data Synthesis

Subgroup Analyses

Subgroup analyses will be used to evaluate the relationship between moderators and the magnitude of the effect size. Given the subgroup analyses, the average effect size for a given level of the moderator can be estimated and group differences can also be tested. These analyses provide information about the influence of both continuous and categorical moderators and the conditions under which PAL may (or may not) be effective. Broad claims regarding the effectiveness of PAL in different contexts, such as ethnicity, region, and university type, offer valuable insights for educators implementing PAL in their courses and institutions. These claims provide context for understanding when and for whom PAL may be effective.

Multiple moderator analyses computed using mixed effects (aka random effects) models, will be conducted. Continuous moderators will be group-mean centered (i.e., centered within the study) prior to analyses, and categorical moderators will be dummy coded and uncentered so that the intercept reflects the outcome when the moderator is the category coded as zero. Moderator effects in the mixed-effects meta-regression models will be tested using the Knapp and Hartung method, which controls Type I error rates and is recommended as the most suitable meta-analytic option (Viechtbauer et al., 2015). Also, because there are multiple moderators and the possibility of interactions between moderators, *meta-CART* (R package) will be used. Meta-CART combines classification and regression trees (CART) to identify interactions in meta-analyses and then tests the significance of the moderator effects with subgroup meta-analysis (Li et al., 2019). We will likely not be able to account for all variability of methods of implementation and the potential interaction with other teaching strategies—nor the extent that other teaching strategies may be being implemented. This is inevitable in a meta-analysis given reliance on author reported information. To address this, the included moderators address different ways of implementing PAL, and the review team will maintain flexibility in coding to add appropriate PAL-related moderators.

Sensitivity Analyses

Sensitivity analyses will be conducted to examine the extent to which the results are robust to outlying effect sizes. Visual means, such as forest plots, will be used to aid in identifying potential outliers for possible exclusion from the analytic sample.

Preliminary Timeframe

The literature search was conducted in Summer 2023. Abstract and title screening, full-text eligibility screening, and data extraction coding training is in progress. The approximate date for completion of the study is December 2024.

Acknowledgements

This material is based upon work supported by the National Science Foundation (NSF) under grant number 2236074. Any opinions, findings, and conclusions or recommendations expressed

in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

- Achieving the Dream. (2021a). *Engaging students with adaptive courseware in gateway math courses*. Achieving the Dream. Retrieved July 5 from <https://achievingthedream.org/every-learner-everywhere-case-study-adaptive-learning-technology-in-mathematics/>
- Achieving the Dream. (2021b). *Engaging students with adaptive courseware in gateway math courses*. Achieving the Dream. Retrieved July 5 from <https://achievingthedream.org/every-learner-everywhere-case-study-adaptive-learning-technology-in-mathematics/>
- Achieving the Dream. (2021c). *Integrating adaptive courseware as part of a comprehensive redesign of a gateway math course*. Achieving the Dream. Retrieved July 5 from <https://achievingthedream.org/every-learner-everywhere-case-study-miami-dade-college/>
- Achieving the Dream. (2021d). *Integrating adaptive courseware into broader efforts to improve teaching and learning*. Achieving the Dream. Retrieved July 5 from <https://achievingthedream.org/every-learner-everywhere-case-study-indian-river-state-college/>
- Achieving the Dream. (2021e). *Statistical significance: Implementing adaptive courseware in gateway math and business courses*. Achieving the Dream. Retrieved July 5 from <https://achievingthedream.org/every-learner-everywhere-case-study-lorain-county-community-college/>
- Achieving the Dream. (2021f). *Supporting greater integration of adaptive courseware in gateway courses*. Achieving the Dream. Retrieved July 5 from <https://achievingthedream.org/every-learner-everywhere-case-study-broward-college/>
- Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2017). Instruction based on adaptive learning technologies. In R. E. Mayer & P. Alexander (Eds.), *Handbook of Research on Learning and Instruction* (2 ed., pp. 552-560).
- Aroyo, L., Dolog, P., Houben, G.-J., Kravcik, M., Naeve, A., Nilsson, M., & Wild, F. (2006). Interoperability in personalized adaptive learning [Article]. *Journal of Educational Technology & Society*, 9(2), 4-18. <https://search.ebscohost.com/login.aspx?direct=true&db=eue&AN=85866413&site=eds-live&scope=site&custid=current&groupid=main&authtype=shib>
- Bailey, T., Bashford, J., Boatman, A., Squires, L., Weiss, M., Doyle, W., Valentine, J. C., LaSota, R., Polanin, J. R., Spinney, E., Wilson, W., Yeide, M., & Young, S. H. (2016). *Strategies for postsecondary students in developmental education—A practice guide for college and university administrators, advisors, and faculty*. <https://ies.ed.gov/ncee/wwc/PracticeGuide/23>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research synthesis methods*, 1, 97-111. <https://doi.org/10.1002/jrsm.12>
- Bryk, A. S., & Treisman, U. (2010). Make math a gateway, not a gatekeeper. *Chronicle of Higher Education*(32). <https://search.ebscohost.com/login.aspx?direct=true&db=edsggo&AN=edsgcl.224347923&site=eds-live&scope=site&custid=current&groupid=main&authtype=shib>
- Center for Distributed Learning. *Personalized adaptive learning*. Retrieved January 23 from <https://cdl.ucf.edu/teach/pal/#:~:text=Personalized%20adaptive%20learning%20is%20a>

- [%20software%20platform%20approach%20that%20provides,%2C%20skills%2C%20and%20learning%20needs.](#)
- Centre for Reviews and Dissemination. (2009). *Systematic reviews: CRD's guidance for undertaking reviews in health care*. Centre for Reviews and Dissemination. https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf
- Chan, A. W., Song, F., & Vickers, A. (2014). Increasing value and reducing waste: Addressing inaccessible research. *Lancet*, 383, 257-266. [https://doi.org/10.1016/S0140-6736\(13\)62296-5](https://doi.org/10.1016/S0140-6736(13)62296-5)
- Chang, M. J., Cerna, O., Han, J., & Saenz, V. (2008). The contradictory roles of institutional status in retaining underrepresented minorities in biomedical and behavioral science majors. *Review of Higher Education*, 31(4), 433-464. <https://doi.org/10.1353/rhe.0.0011>
- Chen, X. (2013). *STEM attrition: College students' paths into and out of STEM fields*.
- Chen, X. (2016). *Remedial coursetaking at U.S. public 2- and 4-year institutions: Scope, experiences, and outcomes* (NCES 2016-405).
- Chen, X., & Soldner, M. (2013). *STEM Attrition: College Students' Paths Into and Out of STEM Fields* (NCES 2014-001).
- Cochrane Collaboration. (2017). *Cochrane Effective Practice and Organisation of Care (EPOC): EPOC resources for review authors*. Retrieved November 9 from epoc.cochrane.org/resources/epoc-resources-review-authors
- CRAFTY. (2007). *College algebra guidelines*. <https://www.maa.org/sites/default/files/pdf/CUPM/crafty/CRAFTY-Coll-Alg-Guidelines.pdf>
- Crisp, G., Nora, A., & Taggart, A. (2009). Student characteristics, pre-college, college, and environmental factors as predictors of majoring in and Earning a STEM degree: An analysis of students attending a Hispanic Serving Institution. *American educational research journal*, 46(4), 924-942. <https://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=45668049&site=eds-live&scope=site&custid=current&groupid=main&authtype=shib>
- Cuijpers, P., Weitz, E., Cristea, I. A., & Twisk, J. (2017). Pre-post effect sizes should be avoided in meta-analyses. *Epidemiology and Psychiatric Sciences*, 26(4), 364-368. <https://doi.org/10.1017/S2045796016000809>
- § § 6301 (2015). <https://www.congress.gov/bill/114th-congress/senate-bill/1177>
- Fox, K., Khedkar, N., Bryant, G., NeJame, L., Dorn, H., & Nguyen, A. (2021). *Time for class 2021: The state of digital learning and courseware adoption*. Tyton Partners. Retrieved July 5 from <https://www.everylearnereverywhere.org/wp-content/uploads/Time-for-Class-2021.pdf>
- Frandsen, T. F., Gildberg, F. A., & Tingleff, E. B. (2019). Searching for qualitative health research required several databases and alternative search strategies: A study of coverage in bibliographic databases. *Journal of Clinical Epidemiology*, 114, 118-124. <https://doi.org/10.1016/j.jclinepi.2019.06.013>
- Ganter, S., & Haver, E. (Eds.). (2011). *Partner discipline recommendations for introductory college mathematics and the implications for college algebra*. Mathematical Association of America. <https://www.maa.org/sites/default/files/pdf/CUPM/crafty/introreport.pdf>.
- Gobel, S., Wendel, V., Ritter, C., & Steinmetz, R. (2010). Personalized, adaptive digital educational games using narrative game-based learning objects. International Conference on E-learning and Games, Edutainment 2010, Changchun, China.

- Gordon, S. P., & Gordon, F. S. (2021). Incorporating statistics into college algebra. *International Journal for Technology in Mathematics Education*, 28(1), 47.
https://doi.org/10.1564/tme_v28.1.04
- Hagedorn, L. S., & DuBray, D. (2010). Math and science success and nonsuccess: Journeys within the community college. *Journal of women and minorities in science and engineering*, 16(1), 31-50. <https://doi.org/10.1615/JWomenMinorScienEng.v16.i1.30>
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107-128.
<https://doi.org/10.2307/1164588>
- Hey, T., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: Data-intensive scientific discovery*. Microsoft Research.
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2022). *Cochrane handbook for systematic reviews of interventions*.
<https://training.cochrane.org/handbook#how-to-access>
- Howlin, C. (2014). *Realizeit at the University of Central Florida: Results from initial trials of Realizeit at the University of Central Florida*. RealizeIt.
- Hurtado, S., Eagan, K., & Chang, M. (2010). *Degrees of success: Bachelor's degree completion rates among initial STEM majors*. <https://heri.ucla.edu/nih/downloads/2010-Degrees-of-Success.pdf>
- Johnson, D., & Samora, D. (2016). The potential transformation of higher education through computer-based adaptive learning systems. *Global Education Journal*, 2016(1), 1-17.
<https://search.ebscohost.com/login.aspx?direct=true&db=eue&AN=112407351&site=eds-live&scope=site&custid=current&groupid=main&authtype=shib>
- Jones, L. E., & Casey, M. C. (2015). *Personalized learning: Policy and practice recommendations for meeting the needs of students with disabilities*. National Center for Learning Disabilities. Retrieved December 8 from <https://www.nclld.org/wp-content/uploads/2016/04/Personalized-Learning.WebReady.pdf>
- Joo, J., & Spies, R. R. (2019). Aligning many campuses and instructors around a common adaptive learning courseware in introductory statistics: Lessons from a multi-year pilot in Maryland. <https://sr.ithaka.org/publications/adaptive-learning-courseware-introductory-statistics/>
- Kersaint, G., Dogbey, J., Barber, J., & Kephart, D. (2011). The effect of access to an online tutorial service on college algebra student outcomes. *Mentoring & Tutoring*, 19(1), 25-44. <https://doi.org/10.1080/13611267.2011.543568>
- Li, T., Higgins, J. P. T., & Deeks, J. J. (2023). Chapter 5: Collecting data. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions (version 6.4)*. Cochrane.
<https://training.cochrane.org/handbook>
- Li, X., Dusseldorp, E., & Meulman, J. J. (2019). A flexible approach to identify interaction effects between moderators in meta-analysis. *Research synthesis methods*, 10(1), 134-152. <https://doi.org/10.1002/jrsm.1334>
- Lin, C. F., Yeh, Y.-c., Hung, Y. H., & Chang, R. I. (2013). Data mining for providing a personalized learning path in creativity: An application of decision trees [Article]. *Computers & Education*, 68, 199-210. <https://doi.org/10.1016/j.compedu.2013.05.009>
- Lin, L., & Chu, H. (2018). Quantifying publication bias in meta-analysis. *Biometrics*, 74(3), 785-794. <https://doi.org/10.1111/biom.12817>

- Liu, M., McKelroy, E., Corliss, S., & Carrigan, J. (2017). Investigating the effect of an adaptive learning intervention on students' learning [Article]. *Educational Technology Research & Development*, 65(6), 1605-1625. <https://doi.org/10.1007/s11423-017-9542-1>
- McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016a). *PRESS Peer Review of Electronic Search Strategies: 2015 guideline explanation and elaboration*. https://www.cadth.ca/sites/default/files/pdf/CP0015_PRESS_Update_Report_2016.pdf
- McGowan, J., Sampson, M., Salzwedel, D. M., Cogo, E., Foerster, V., & Lefebvre, C. (2016b). PRESS Peer Review of Electronic Search Strategies: 2015 guideline statement. *Journal of Clinical Epidemiology*, 75, 40-46. <https://doi.org/10.1016/j.jclinepi.2016.01.021>
- Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gotzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., & Altman, D. G. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ: British Medical Journal*, 340.
- National Academies of Sciences, n., nd Medicine. (2016). Barriers and opportunities for 2-year and 4-year STEM degrees: Systemic change to support diverse student pathways. In *Committee on barriers and opportunities in completing 2-year and 4-year STEM degrees, board on science education, and board on higher education and the workforce*. The National Academies Press.
- National Center for Science and Engineering Statistics. (2021). *Women, minorities, and persons with disabilities in science and engineering* (NSF 21-321). <https://nces.nsf.gov/pubs/nsf21321>
- Page, M. J., Higgins, J. P. T., & Sterne, J. A. C. (2020, September 2020). *Chapter 13: Assessing risk of bias due to missing results in a synthesis*. Cochrane. Retrieved November 17, 2020 from https://training.cochrane.org/handbook/current/chapter-13#_Ref514839621
- Page, M. J., Moher, D., 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., . . . McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *British Medical Journal*, 372, n160. <https://doi.org/10.1136/bmj.n160>
- Peng, H., Ma, S., & Spector, J. M. (2019). Personalized adaptive learning: an emerging pedagogical approach enabled by a smart learning environment [article in journal/newspaper]. <https://doi.org/10.1186/s40561-019-0089-y>
- Polanin, J. R., Pigott, T. D., Espelage, D. L., & Grotzinger, J. K. (2019). Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. *Research synthesis methods*, 10(3), 330-342. <https://doi.org/10.1002/jrsm.1354>
- President's Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering and mathematics*. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf
- Rech, J. F., & Harrington, J. (2000). Algebra as a gatekeeper: A descriptive study at an urban university [Article]. *Journal of African American Men*, 4(4), 63-71. <https://doi.org/10.1007/s12111-000-1022-7>

- Relevo, R., & Paynter, R. (2012). *Peer review of search strategies*.
<https://www.ncbi.nlm.nih.gov/books/NBK98353/>
- Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, 10(1), 39-39.
<https://doi.org/10.1186/s13643-020-01542-z>
- Richardson, W. S., Wilson, M. C., Nishikawa, J., & Hayward, R. S. (1995). The well-built clinical question: A key to evidence-based decisions. *ACP journal club*, 123(3), A12-A13. <https://doi.org/10.7326/ACPJC-1995-123-3-A12>
- Rutschow, E. Z., Cormier, M. S., Dukes, D., & Cruz Zamora, D. E. (2019). *The changing landscape of developmental education practices: Findings from a national survey and interviews with postsecondary institutions*.
<https://ccrc.tc.columbia.edu/publications/changing-landscape-developmental-education-practices.html#:~:text=Opinion-.The%20Changing%20Landscape%20of%20Developmental%20Education%20Practices%3A%20Findings%20From%20a,and%20Interviews%20with%20Postsecondary%20Institutions&text=Research%20suggests%20that%20far%20more,a%20barrier%20to%20students%20success.>
- Saxe, K., & Braddy, L. (2015). *A common vision for undergraduate mathematical sciences programs in 2025*. Mathematical Association of America.
- Simon-Campbell, E. L., & Phelan, J. (2016). Effectiveness of an adaptive quizzing system as an institutional-wide strategy to improve student learning and retention. *Nurse Educator*, 41(5), 246-251. <https://doi.org/10.1097/NNE.0000000000000258>
- Spry, C., Mierzwinski-Urban, M., & Rabb, D. (2013). Peer review of literature search strategies: does it make a difference? Cochrane Colloquium, Quebec City, Canada.
- Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, 54, 1046-1055.
[https://doi.org/10.1016/s0895-4356\(01\)00377-8](https://doi.org/10.1016/s0895-4356(01)00377-8)
- Stevenson, C., & Lawlor, D. A. (2004). Searching multiple databases for systematic reviews: Added value or diminishing returns? *Complementary therapies in medicine*, 12(4), 228-232. <https://doi.org/10.1016/j.ctim.2004.09.003>
- Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. In (Vol. 40, pp. 604-634).
- Toven-Lindsey, B., Levis-Fitzgerald, M., Barber, P. H., & Hasson, T. (2015). Increasing persistence in undergraduate science majors: A model for institutional support of underrepresented students. *CBE - Life Sciences Education*, 14(2), 1-12.
<https://doi.org/10.1187/cbe.14-05-0082>
- U.S. Department of Education Office of Educational Technology. (2017). *Reimagining the role of technology in education: 2017 national education technology plan update*.
<https://tech.ed.gov/files/2017/01/NETP17.pdf>
- Viechtbauer, W., López-López, J. A., Sánchez-Meca, J., & Marín-Martínez, F. (2015). A comparison of procedures to test for moderators in mixed-effects meta-regression models. *Psychological methods*, 20(3), 360-374. <https://doi.org/10.1037/met0000023>
- Waffenschmidt, S., Knelangen, M., Sieben, W., Bühn, S., & Pieper, D. (2019). Single screening versus conventional double screening for study selection in systematic reviews: A

- methodological systematic review. *BMC Medical Research Methodology*, 19(1), 132-132. <https://doi.org/10.1186/s12874-019-0782-0>
- Wallace, S., Daly, C., Campbell, M., Cody, J., Grant, A., Vale, L., Donaldson, C., Khan, I., Lawrence, P., & MacLeod, A. (1997). *After MEDLINE? Dividend from other potential sources of randomised controlled trials* Second International Conference Scientific Basis of Health Services & Fifth Annual Cochrane Colloquium, Amsterdam, Netherlands.
- Weiss, M. J., & Headlam, C. (2018). *A randomized controlled trial of a modularized, computer-assisted self-paced approach to developmental math*. MDRC.
- Wheeler, S. W., & Bray, N. (2017). Effective evaluation of developmental education: A mathematics example. *Journal of Developmental Education*, 41(1), 10-17. <https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1192548&site=eds-live&scope=site&custid=current&groupid=main&authtype=shib>
- Xie, H., Chu, H.-C., Hwang, G.-J., & Wang, C.-C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2019 [Article]. *Computers & Education*, 140. <https://doi.org/10.1016/j.compedu.2019.103599>
- Yarnall, L., Means, B., & Wetzel, T. (2016). Lessons learned from early implementations of adaptive courseware. <https://www.sri.com/publication/education-learning-pubs/digital-learning-pubs/lessons-learned-from-early-implementations-of-adaptive-courseware/>
- Zhang, L., Basham, J. D., & Yang, S. (2020). Understanding the implementation of personalized learning: A research synthesis. *Educational Research Review*, 31. <https://doi.org/10.1016/j.edurev.2020.100339>