IES/NSF Pipeline-of-Evidence Protocol as Explanation for Successes and Failures of Gates Foundation Funded Initiatives

Barbara R. Schirmer *
Walden University, USA

Alison S. Lockman
Walden University, USA

Todd N. Schirmer
Saybrook University, USA

Abstract: This study was designed to investigate the applicability of the IES/NSF pipeline-of-evidence protocol in ascertaining why two notable educational initiatives spearheaded and financially supported by the Bill and Melinda Gates Foundation achieved or not the goal of improved academic outcomes for K-12 public school students. Our interest was not whether there is a sufficient body of high quality research evidence to support the two initiatives but whether the research considered by the Gates Foundation established the likelihood that the initiatives would be successful and worth the decision to dedicate substantial funding, time, and effort required for each versus the many competing programs seeking sponsorship. We found that in the case of Intensive Partnerships for Effective Teaching, foundational, efficacy, and effectiveness research were absent and the Gates foundation discontinued grant support because the initiative had not achieved the goal of improved high school graduation and college attendance among low-income minority students. In the case of Early College High School, we found empirical evidence was manifest at all but the effectiveness stage of the pipeline and the initiative continued to receive funding. Our findings support the importance of widening the net of methodologies that constitute a framework for elements needed to make predictions of effectiveness for any given intervention before investing in scale-up initiatives and the need for private foundations to be transparent in their decision-making process to enable others to scrutinize the research that informs the design of initiatives.

Keywords: *Evidence-based practices, K-12 education initiatives, early college high schools, teacher evaluation, dual enrollment.*

Introduction

The Bill and Melinda Gates Foundation has provided $41.3 billion in grant support to numerous initiatives designed to improve the lives of vulnerable populations since its founding in 2000. The goal of the K-12 initiatives in the U.S. has been to ensure that all students receive an education that enables them to pursue post-secondary education, attain successful careers, and achieve social mobility and personal fulfillment. Given the substantial size of grant funding, publicity surrounding the initiatives selected by the foundation, and controversy about the influence of a private foundation on trends in K-12 education, it is important to identify the bases for decisions by the foundation for its funding priorities and whether the initiatives reflect scientific evidence indicating the likelihood of successful outcomes. Since the first K-12 grant-funded initiative in 2006, outcomes have been mixed. Given the influence of Gates funding on K-12 education in the U.S., it is important to determine whether flaws in the selection process or design of the initiatives offer reasons for successes and failures.

The process used by the Gates Foundation for deciding which initiatives to fund from among the many ideas that are brought to the foundation's attention is not published. However, it is possible to retroactively examine the prior research reported to have led to the selection and design of initiatives based on Gates foundation-published evaluation reports. For such an examination, we used the framework of line of research inquiry, which connotes building a body of knowledge from study to study. The term implies that knowledge derived from research proceeds from observing and describing phenomena, to uncovering the links between phenomena, and then to influencing phenomena in order to generate particular outcomes. The principle of converging evidence has been proposed as a means for drawing on the findings from studies employing different designs in order to conclude whether a practice is research-based and ready

---

* **Corresponding author:**
Barbara R. Schirmer, PO Box 6094, Monroe Township, NJ, USA. ✉ barbara.schirmer@mail.waldenu.edu

to be scaled up. Given this perspective, we applied the protocol developed jointly by two U.S. federal agencies as the most appropriate for reviewing the Gates-funded initiatives.

The U.S. Institute of Education Sciences and the National Science Foundation (IES/NSF) issued common guidelines for education research and development in 2013. Their purpose was "to identify the spectrum of study types that contribute to development and testing of interventions and strategies and to specify expectations for the contributions of each type of study" (p. 8). IES/NSF described relevant educational research as forming a pipeline-of-evidence that contributes to the accumulation of empirical evidence and development of theoretical models. Unlike previous efforts to determine which studies provide sufficient evidence to identify an educational practice as research-based (e.g., Cooper, 2010; Council for Exceptional Children, 2014; Gersten et al., 2005; Kazdin, 2011; What Works Clearinghouse, 2017), the IES/NSF guidelines provide a protocol by which the use of particular methodological designs in a line of research inquiry provides evidence for each successive step in the process of bringing any given instructional intervention into practice.

The current study was designed to investigate the applicability of the IES/NSF pipeline-of-evidence protocol in exploring why two notable educational initiatives spearheaded and financially supported by the Bill and Melinda Gates Foundation achieved or not the goal of improved academic outcomes for K-12 public school students in the U.S.

Methodology

Design and Data Sources

We used a qualitative methodological design to explore the narrative data on two Gates Foundation initiatives. All data sources were cited in Gates reports about each initiative. That is, we did not identify sources that the Gates Foundation might have considered when deciding on a given initiative but, rather, we only reviewed the sources that Gates did consider. Our goal was to examine the empirical evidence for the two Gates initiatives and map this evidence to the steps in the IES/NSF pipeline of evidence.

Procedure

Our first step was to select the initiatives. We chose Intensive Partnerships for Effective Teaching and Early College High School initiatives for two reasons. One is that the initiatives had been in effect for enough time to determine if they were sufficiently successful for the Gates Foundation to continue support beyond the initial phase of patronage. The second reason is that each had been highly touted as having the potential to be a disruptive innovation in education at their launch.

The Intensive Partnerships for Effective Teaching, which we will refer to as Teacher Evaluation, was begun in 2009-2010 to improve high school graduation and college attendance rates among low-income minority students by improving the effectiveness of their teachers. Each educational site developed a measure of teaching effectiveness that included the teacher's contribution to growth in student achievement and assessment of teaching based on classroom observations. Each site was then expected to use assessment data to improve recruitment and hiring practices, adjust placement and transfer of teachers to ensure that low-income minority students received the most effective teachers, reform tenure and dismissal policies so that the most effective teachers were advanced and ineffective teachers were removed, link professional development to teachers' weaknesses, and add financial and promotion incentives to retain the most effective teachers.

The Early College High School initiative, which we will refer to as Early College, was launched in 2002 to increase the opportunity for underserved students to earn an associate's degree or up to two years of college credits applicable to a bachelor's degree while still in high school and avoid the non-degree-applicable developmental courses in reading and math that are commonly required for underprepared students during their early college semesters. Early Colleges were expected to be created through a partnership between high schools and postsecondary institutions located in proximity to each other. The Early College schools and higher education partners must be committed to serving underrepresented students and accountable for student success, establish an integrated academic program that enabled students to earn 1-2 years of transferable college credit so that the student can apply these credits toward college completion, and offer a comprehensive support system to help develop students' academic, social, and behavioral skills necessary for college success.

Our second step involved collecting and examining the prior research that led to the design of the two initiatives. We found the empirical evidence in program evaluation reports conducted by independent evaluators at the American Institutes for Research (Berger et al., 2013; Stecher et al., 2018). The data sources included the two program evaluation reports and the research studies cited within each report. We made copies of each report and research study for data analysis.

For the third step, we mapped this evidence to the following steps in the IES/NSF pipeline-of-evidence protocol:

Research Type 1: Foundational involves studies that provide foundational knowledge of teaching and learning, develop and refine theory, and examine phenomena in the absence of a direct link to educational outcomes.

Research Type 2: Early Stage/Exploratory involves studies that examine the connections or relationships among constructs that may result in the development of a new intervention.

Research Type 3: Design and Development involves studies that draw on theory and empirical evidence in designing an intervention and testing individual components.

Research Type 4: Efficacy involves studies that test the intervention under ideal circumstances.

Research Type 5: Effectiveness involves studies that test the intervention under typical circumstances.

Research Type 6: Scale-up involves studies that test the intervention under typical circumstances but in a wide range of contexts and populations.

Data Analysis

We began data analysis by organizing the data sources into meaningful units of analysis (Creswell & Creswell, 2018). We used the IES/NSF protocol as the a priori scheme for deductive coding. Each author first reviewed the two program evaluation reports and then reviewed the sources cited in each of these reports that were identified as supporting the decision to launch each initiative. Each of us then explored these documents by first reading through them for an overall sense of the content and then writing notes for each during careful readings (Saldaña, 2016). The a priori scheme enabled us to focus on the characteristics of foundational, early stage/exploratory, design and development, efficacy, effectiveness, and scale-up research and match these characteristics to the design of each study as appropriate for following the requirements of qualitative deductive analysis (Ravitch & Carl, 2016).

To address potential issues of trustworthiness, we employed strategies to assure credibility, transferability, dependability, and confirmability (Patton, 2014). We engaged in researcher reflexivity by considering our assumptions, beliefs, values, and biases that could influence our interpretation of the data. We kept an audit trail to provide tracking of questions, insights, and decisions during data analysis. We independently analyzed the corpus of data and then compared our findings, discussed discrepancies, and reached consensus for each study.

Findings

Teacher Evaluation

We found no discussion of research that provided foundational knowledge (Research Type 1) though each Teacher Evaluation site was expected to follow a theory of action by specifying steps in implementing changes to the site's approach to teacher evaluation and a plan for evaluating outcomes. The constructs of the theory of action included (a) a valid measure of teacher effectiveness, (b) staffing policies that linked the measure of teacher effectiveness to hiring, placement, retention, and dismissal decisions, (c) customized professional development to meet the needs of individual teachers, and (d) compensation and career ladder policies for teachers (Stecher et al., 2018).

Research on the relationships between aspects of teacher performance and student achievement were identified as early stage or exploratory research used in identifying potential components for the Teacher Evaluation initiative (Research Type 2). Findings from this body of research indicated that though teachers have a significant influence on student achievement, variables of teacher experience and graduate education (Hanushek, 1971; Rivkin et al., 2005) and compensation incentives (Fryer, 2011; Springer et al., 2012) did not contribute to student achievement growth.

Three lines of research influenced the design and development of the teacher assessments used in the Teacher Evaluation initiative (Research Type 3). One line was the research on the benefits of multiple measures of teacher effectiveness versus any single measure as predictive of student achievement growth. Findings showed that multiple measures of teacher performance, with or without incentives, improved student learning outcomes (Dee & Wyckoff, 2015; Taylor & Tyler, 2012). The second line of research involved the acceptability of evaluation measures to the teachers. Findings showed that teachers felt positive about evaluations involving observations but skeptical about including student growth measures (Jiang et al., 2015). The third line of research inquiry involved testing the reliability and validity of feedback from student surveys (Kane et al., 2010) and teacher observation protocols (Kane & Staiger, 2012). The qualities of student surveys and observation protocols found to be effective were then recommended to the Teacher Evaluation sites funded by the Gates Foundation.

We found no studies that involved testing the intervention under ideal circumstances (Research Type 4) or typical circumstances (Research Type 5) prior to scaling up the intervention in a wide range of contexts and populations (Research Type 6) for the Teacher Evaluation initiative funded by the Gates Foundation. According to Stecher et al. (2018), the decision to launch the Teacher Evaluation initiative was based on the premise that high-quality measures of teacher effectiveness would improve instruction and that a high-quality measure should include three sources of information: student achievement growth, observation of teacher performance, and student feedback.

According to the evaluation report, the Teacher Evaluation initiative did not lead to gains in student achievement, graduate rates, or post-school outcomes (Stecher et al., 2018). Results showed that most teachers were rated in the higher categories of effectiveness between the 2011-2012 and 2013-2014 school years, and the lowest category contained 2% or less by 2014. It was also found that sites had difficulty recruiting effective teachers to high-need schools, just 1% of teachers were dismissed for poor performance, sites were unable to individualize professional development based on teacher evaluations, modifications to compensation policies to reward effective teaching did not lead to more effective teachers being placed in higher need schools, the salary gap between more and less effective teachers did not change, and new career ladder policies did not incorporate the steps and increased responsibilities that resulted in increased retention of the most effective teachers.

Early College

We found that the Early College initiative was based on three foundational principles: high school rigor to build content knowledge and learning skills, relevance of the high school curriculum for making real-world connections for students, and relationships with instructors and peers during high school to support engagement and achievement (Research Type 1). The Early College initiative incorporated five activities to address these principles (Tierney et al., 2009). For students to be academically prepared for college by grade 12, the Early College schools were expected to (a) incorporate college preparatory courses by grade 9 and (b) provide regular assessment information to students to assist them in attending to academic weaknesses. To address the need for students to negotiate the many tasks during high school that lead to college enrollment about which their families may not be knowledgeable, the Early College schools were also expected to (c) ensure access to adults and peers who support college aspirations, (d) assist students with the tasks of taking college entrance exams, identifying potential colleges, and completing college and financial aid applications, and (e) assist students and their families with understanding the financial obligations of college and applying for financial aid.

Research on the relationships among several constructs were represented in the Early College initiative (Research Type 2). One set of constructs were investigated in studies on the relationship between college degree attainment and wages, which showed that each level of postsecondary education adds a significant amount to lifetime earnings (Carnevale et al., 2011). Another set involved the relationship between college degree attainment and demographic characteristics, which demonstrated significantly lower degree attainment for individuals from underrepresented populations and disadvantaged families (Aud et al., 2010). These findings led to the goal of two years of college credits during Early College and the focus on recruiting students from minority and disadvantaged families.

Research on dual enrollment programs, which involve college courses taken by high school students on a high school or college campus, offers evidence of designing the intervention and testing components (Research Type 3). These dual credit courses have been found to promote college enrollment through access to college-level academic and technical courses while in high school (Kleiner & Lewis, 2005) and to be associated with greater high school persistence and college degree completion (Karp et al., 2007; U.S. Department of Education, 2017). These findings provided support for incorporating college-level coursework in Early Colleges.

The research on dual enrollment programs also reflects efficacy research for the Early College component of college-level coursework (Research Type 4). Given the positive outcomes of dual credit courses, the initiative included free degree-applicable college courses as a key component. However, as it had also been found that the informal recruitment common in dual enrollment programs resulted in a lack of diversity among the students (Hughes et al., 2006), the design of the Early College initiative incorporated targeted recruitment of diverse students.

We found no studies that involved testing the initiative within one or a few highly controlled settings (Research Type 5) prior to scaling up by testing the intervention in a wide range of contexts and populations (Research Type 6) that were funded by the Gates Foundation for the Early College initiative (Berger et al., 2013).

According to the evaluation report, students were significantly more likely to graduate high school, enroll in college, and earn a college degree in Early College High Schools than students in comparison high schools. The comparison students reflected similar demographics to the Early College students but attended larger high schools with fewer academic supports and less direct attention to college readiness than the Early College High Schools. Results did not generally differ for subgroups and when they did, outcomes were stronger for female than male, minority than non-minority, lower income than higher income, and higher middle school achievement than lower achieving students (Berger et al., 2013). (See the appendix for a summary of the findings)

Discussion

We began this study with the question of whether the IES/NSF pipeline-of-evidence protocol is applicable in ascertaining why two notable educational initiatives spearheaded and financially supported by the Bill and Melinda Gates Foundation achieved or not the goal of improved academic outcomes for K-12 public school students. We applied the IES/NSF pipeline-of-evidence guidelines to assess whether the Teacher Evaluation and Early College initiatives were based on a research base for effectiveness that emerged from an accumulation of empirical evidence and

identification of conceptual or theoretical frameworks. Our interest was not whether there is a sufficient body of high quality research evidence to support the two initiatives but whether the research considered by the Gates Foundation established the likelihood that the initiatives would be successful and worth the decision to dedicate substantial funding, time, and effort required for each versus the many competing programs seeking sponsorship.

We found that in the case of Teacher Evaluation, foundational, efficacy, and effectiveness research were absent. The Gates foundation discontinued grant support because the initiative had not accomplished the goal of improved high school graduation and college attendance rates among low-income minority students. In the case of Early College, we found empirical evidence was manifest at all but the effectiveness stage of the pipeline. The Gates Foundation continued to provide support to the initiative through increasing numbers of Early Colleges (Bill and Melinda Gates Foundation, 2019).

In a prior study of applying the IES/NSF pipeline-of-evidence protocol, Schirmer et al. (2016) applied the protocol to assess whether instructional practices touted as having a research base for effectiveness have emerged from an accumulation of empirical evidence and identification of conceptual or theoretical frameworks. Results indicated that the protocol offered a productive approach to identifying evidence-based practices because it takes into account the role of methodological designs in lines of research inquiry. The findings of the present study align with this previous study and confirm conclusions that external funding agencies and foundations should widen the net of methodologies that constitute a framework for elements needed to make predictions of effectiveness for any given educational intervention or program.

Conclusion

Our approach to examining two highly touted K-12 initiatives that were implemented with funding from the Gates Foundation offers an approach to identifying the likelihood of success for any new educational practice. Rather than leaping from good idea to scale-up interventions, our findings underlie the importance of a line of research inquiry that provides researchers, practitioners, policymakers, federal agencies, and private foundations with a pipeline-of-evidence for designing and implementing interventions with the likelihood of effectiveness in diverse educational environments. However, the approach is dependent on transparency of the body of research considered when a new intervention is launched. Despite the role of private foundations in the U.S. for shaping K-12 educational practices through grant funding, their decision-making process is proprietary and not available for those outside the foundation to scrutinize and critique the research that informed the design of initiatives. Our findings support the importance of widening the net of methodologies that constitute a framework for elements needed to make predictions of effectiveness for any given intervention before investing in scale-up initiatives.

Suggestions

Our findings suggest that funding agencies must not only weigh the quality of the research that authors use to make the case for the importance of their intervention but also the methodological designs used in prior research. The history of the research should demonstrate that phenomena have been explored, key variables and their relationship to each other identified, and the intervention has been tested in a controlled environment. Only then should a decision be made that an intervention is ready to be tested in wider settings that offer the potential for greater educational impact.

Given the importance of research at all stages in the pipeline, our findings also suggest that funding for educational research should be aimed as much at foundational, exploratory, design, and efficacy as at effectiveness and scale-up studies. By funding studies that employ methodologies at each stage in the protocol, the likelihood will be greater that experimental investigations of instructional interventions at the latter stages of the protocol will show evidence of effectiveness in improving outcomes because earlier research led to the development of the intervention.

Limitations

The study is limited by the lack of transparency regarding the process used by the Gates Foundation for selecting initiatives. In order to determine whether the research base should have warranted the levels of funding and widespread publicity that the initiatives received, we had to work inductively to identify whether research findings at each step in the pipeline-of-evidence protocol were considered. It may be that the gaps we found do not reflect the full corpus of research considered by the Gates Foundation in the initiative selection process.

Our findings are also limited by the two initiatives we examined. The Gates Foundation has funded numerous projects aimed at improving K-12 education. We selected just two for our study. It may be that others implemented at about the same time as the two we investigated would reveal different evidence bases before launch and different trajectories subsequently. It may also be that the process for selecting initiatives has been modified and that those launched more recently have been selected only after research at each stage in the pipeline indicate scaling up is the logical next step.

We do not suggest that methodology is the only factor that may explain the differential success of the two initiatives. We recognize that whenever an educational intervention is scaled up, there will be intervening variables affecting the outcomes. Further research can address this limitation in our findings by applying the pipeline-of-evidence protocol to

other scaled up interventions to determine if the methodologies in the prior research would have predicted the outcomes.

References

Aud, S., Fox, M. A., & KewalRamani, A. (2010). *Status and trends in the education of racial and ethnic groups*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. https://nces.ed.gov/programs/raceindicators/

Berger, A., Tirl-Bicakci, L., Garet, M., Song, M., Knudson, J., Haxton, C., Zeiser, K., Hoshen, G., Ford, J., Stephan, J., Keating, K., & Cassidy, L. (2013). *Early college, early success: Early College High School initiative impact study.* American Institutes for Research. https://www.air.org/sites/default/files/downloads/report/ECHSI_Impact_Study_Report_Final1_0.pdf

Bill and Melinda Gates Foundation. (2019). *K-12 education.* https://www.gatesfoundation.org/What-We-Do/US-Program/K-12-Education#AREASOFFOCUS

Carnevale, A. P., Rose, S. J., & Cheah, B. (2011). *The college payoff: Education, occupations, lifetime earnings.* Georgetown University, Center on Education and the Workforce. https://cew.georgetown.edu/cew-reports/the-college-payoff/

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Sage.

Council for Exceptional Children. (2014). *Council for Exceptional Children: Standards for evidence-based practices in special education.* https://www.cec.sped.org/~/media/Images/Standards/CEC%20EBP%20Standards%20cover/CECs%20Evidence%20Based%20Practice%20Standards.pdf.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage

Dee, T. W., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management, 34*(2), 267-297. https://doi.org/10.1002/pam.21818

Fryer, R. G. (2011). *Teacher incentives and student achievement: Evidence from New York City Public Schools*. National Bureau of Economic Research, Working Paper 16850. http://www.nber.org/papers/w16850

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149-164. https://doi.org/10.1177/001440290507100202

Hanushek, E. A. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *American Economic Review, 61*(2), 280-288.

Hughes, K. L., Karp, M. M., Fermin, B. J., & Bailey, T. R. (2006). Pathways to college access and success. *Journal for Vocational Special Needs Education, 28*(3), 17-46. https://eric.ed.gov/?id=EJ854373

Institute of Education Sciences and National Science Foundation. (2013). *Common guidelines for education research and development*. Authors. https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf13126

Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher, 44*(2), 105-116.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Bill and Melinda Gates Foundation. https://k12education.gatesfoundation.org/resource/gathering-feedback-on-teaching-combining-high-quality-observations-with-student-surveys-and-achievement-gains-3/

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2010). *Identifying effective classroom practices using student achievement data.* National Bureau of Economic Research. https://www.nber.org/papers/w15803

Karp, M.M., Calcagno, J. C., Hughes, K. L., Jeong, D. W., & Bailey, T. R. (2007). *The postsecondary achievement of participants in dual enrollment: An analysis of student outcomes in two states.* National Research Center for Career and Technical Education, University of Minnesota. https://eric.ed.gov/?id=ED498661

Kazdin, A. E. (2011). *Single-case research design: Methods for clinical and applied settings* (2nd ed.). Oxford University Press.

Kleiner, B., & Lewis, L. (2005). *Dual enrollment of high school students at postsecondary institutions: 2002-03.* U. S. Department of Education, National Center for Education Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2005008

Patton, M. Q. (2015). *Qualitative research and evaluation methods* (4th ed.). Sage.

Ravitch, S. M., & Carl, N. M. (2016). *Qualitative research: Bridging the conceptual, theoretical, and methodological*. Sage.

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools and academic achievement. *Econometrica, 73*(2), *417*-458.

Saldana, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Sage.

Schirmer, B. R., Lockman, A. S., & Schirmer, T. N. (2016). Identifying evidence-based educational practices: Which research designs provide findings that can influence social change? *Journal of Educational Research and Practice, 6*(1), 33-42. https://doi.org/10.5590/jerap.2016.06.1.03

Springer, M. G., Pane, J. F., Le, V-N., McCaffrey, D. F., Burns, S. F., Hamilton, L. S., & Stecher, B. (2012). Team pay for performance: Experimental evidence from the Round Rock Pilot Project on team incentives. *Educational Evaluation and Public Policy Analysis, 34*(4), 367-390. https://doi.org/10.3102/0162373712439094

Stecher, B. M., Holtzman, D. J., Garet, M. S., Hamilton, L. S., Engberg, J., Steiner, E. D., Robyn, A., Baird, M. D., Gutierrez, I. A., Peet, E. D., De Los Reyes, I. B., Fronberg, K., Weinberger, G., Hunter, G. P., & Chambers, J. (2018). *Improving teaching effectiveness: Final report. The Intensive Partners.* American Institutes for Research. https://www.rand.org/pubs/research_reports/RR2242.html

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102*(7), 3628-3651. https://doi.org/10.1257/aer.102.7.3628

Tierney, W. B., Bailey, T. Constantine, J., Finkelstein, N., & Hurd, N. F. (2009). *Helping students navigate the path to college: What high schools can do.* U. S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/PracticeGuide/11

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2017, February). *Dual enrollment programs.* https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_dual_enrollment_022817.pdf

What Works Clearinghouse. (2017). *Procedures and standards handbook version 4.0*. https://ies.ed.gov/ncee/wwc/Handbooks

Appendix. Summary of the Findings

IES/NSF Research Type	Teacher Evaluation	Early College
Research Type 1: Foundational	Theory of action involving steps in implementing changes to teacher evaluation and plan for evaluating outcomes.	Three foundational principles: rigor to build content knowledge and learning skills, real-world connections, and relationships with instructors and peers to build engagement and achievement.
Research Type 2: Early Stage/ Exploratory	Research on relationships between aspects of teacher performance and student achievement.	Research on relationships between college degree attainment and wages, and college degree attainment among minority and low-income students
Research Type 3: Design and Development	Research on the benefits of multiple measures of teacher effectiveness in predicting student achievement growth, acceptability of measures to teachers, and reliability and validity of student surveys and teacher observation protocols	Research on dual enrollment programs and high school persistence, college enrollment, and degree completion
Research Type 4: Efficacy	No studies involved testing the intervention under ideal circumstances.	Research on dual enrollment programs and high school persistence, college enrollment, and degree completion for students from minority and low-income families
Research Type 5: Effectiveness	No studies involved testing the intervention under typical circumstances.	No studies involved testing the intervention under typical circumstances.
Research Type 6: Scale-up	Research on the Teacher Evaluation initiative in multiple sites.	Research on the Early College initiative in multiple sites.