Article

Entrepreneurship Theory and Practice
2022, Vol. 46(6) 1685–1716

DOI: 10.1177/10422587211018184
journals.sagepub.com/home/etp
SAGE

Standards for Evaluating Impact in Entrepreneurship Education Research: Using a Descriptive Validity Framework to Enhance Methodological Rigor and Transparency

Soohyun Yi[1] and Nathalie Duval-Couetil[2]

Abstract

Given calls for more rigorous research able to measure the impact of entrepreneurship education, this study proposes guidelines for enhancing methodological and reporting practices. First, drawing on prior research syntheses, we developed a descriptive validity framework that outlines key elements for rigorous evaluation research. Second, we use this framework to examine 61 quantitative, university-based entrepreneurship education impact studies to identify and describe methodological and reporting practices that are most prevalent. The result is a set of Impact Evaluation Research Standards for entrepreneurship educators and scholars wishing to improve education evaluation research.

Keywords

descriptive validity, program evaluation, entrepreneurship education, systematic review, methodology

Introduction

Entrepreneurship education is a topic of growing interest at universities,[1] based on the belief that it prepares college graduates to thrive in a professional world and economy where energizing business processes, anticipating customer needs, and bringing new ideas to market are highly valued and rewarded. In light of increasing investment and participation in entrepreneurship education, important discussions have emerged around the need for well-designed intervention studies able to measure the impact of entrepreneurship education using methodological rigor and experimental designs (e.g., Martin et al., 2013; Nabi et al., 2017; Rideout & Gray, 2013). While prior work has identified methodological issues from a variety of perspectives, no

[1]Educational Psychology and Leadership, Texas Tech University, Lubbock, TX, , USA
[2]Technology Leadership & Innovation, Purdue University, West Lafayette, IN, USA

Corresponding Author:
Nathalie Duval-Couetil, Technology Leadership & Innovation, Purdue University, West Lafayette, IN 47907, USA.
Email: natduval@purdue.edu

guidelines or standards exist for designing studies or reporting the results of impact evaluation research.

The impact of an educational program is "gauged by the change it produces in outcomes that represent the intended improvements" (Rossi et al., 2019, p. 23). Measuring this change is referred to as *impact evaluation* or *outcome evaluation*, the primary goal of which is to make the least biased inferences by measuring whether the desired outcomes of a program were indeed produced for participants. Complexity in conducting impact evaluation is often due to the influence of factors unrelated to a program. Therefore, rigorous experimental or quasi-experimental research designs are required to disentangle these unintended factors and measure changes in outcomes uniquely explained by an educational program rather than counterfactual conditions (Rossi et al., 2019).

Recent syntheses in entrepreneurship research have identified methodological limitations associated with measuring educational program outcomes, including inferentially weak research designs, few validated assessment instruments, weak statistical power, and little evidence of long-term impact (e.g., Bae et al., 2014; Martin et al., 2013; Nabi et al., 2017; Rideout & Gray, 2013). This research, which reviews published research articles to examine various aspects of impact, notes that many studies in entrepreneurship education are insufficiently designed and reported on. As such, they could not be included in their reviews to adequately synthesize and draw meaningful conclusions about program effectiveness.

Making the details of a research study fully available is referred to as *descriptive validity*, or "the adequacy of the presentation of key features of an evaluation in a research report" (Farrington, 2003). Descriptive validity allows research consumers to make judgments about the quality and generalizability of particular studies based on whether they offer adequate explanations of research design, methods, and results (Gill, 2011). Descriptive validity is strongly related to both internal and external validity, for which transparency is key. This level of detailed reporting is essential for conducting systematic review and meta-analysis studies, and for facilitating discussions about methodological standards and best practices (Farrington, 2003).

In evaluation research, descriptive validity requires acknowledging issues specific to educational practice and assessment that influence methodological rigor. In the case of entrepreneurship education, there are several factors that add complexity to impact evaluation, such as the heterogeneity of program formats (Duval-Couetil et al., 2016; Morris et al., 2013), pedagogical approaches (Fiet, 2001; Neck & Greene, 2011; Pittaway & Cope, 2007), and the breadth and imprecision of outcome measures (Piperopoulos & Dimov, 2015; Pittaway et al., 2009). Despite these obstacles, there is clearly value in striving to employ more rigorous research methods. As educators and scholars, we need objective evidence to help diagnose programmatic issues, justify new or modified program offerings, and verify that educational interventions achieve the desired goals and benefits for targeted populations (Rossi et al., 2019).

To assist in this endeavor, the purposes of this study are to (a) synthesize the methodological issues that prevent entrepreneurship education from meeting rigorous educational evaluation practices and standards and (b) propose standards for impact evaluation research in entrepreneurship education. First, we reviewed guidelines for evaluation research and previous research syntheses in order to create a descriptive validity framework. Second, we used this framework to examine 61 quantitative studies of university-based entrepreneurship education programs, which enabled us to identify and describe methodological practices that are most prevalent in the field. The contribution of this work is a set of Impact Evaluation Research Standards for entrepreneurship educators and scholars wishing to improve assessment practice through more rigorous and reliable research methods and reporting practices.

Literature Review

Challenges Associated With Evaluating the Impact of Entrepreneurship Education

A fundamental role of entrepreneurship education is to teach students how to establish and grow new ventures. Given the important role that entrepreneurship plays in the well-being of society (Wiklund et al., 2019), universities typically invest in entrepreneurship education based on its potential direct contributions to economic development, as well as more indirect measures related to career preparation and personal development. Although creating new startups is at the core of the discipline, many consider the knowledge, skills, and mindset imparted through entrepreneurship education as highly valuable and transferable to a broad set of contexts and professional activities (SÃ¡ & Holt, 2019; Yi & Duval-Couetil, 2018). However, these more indirect outcomes (e.g., knowledge, skills, mindsets) are more difficult to define, assess, and empirically connect to entrepreneurship education when compared to more explicit and fundamental outcomes (e.g., the launch of new ventures).

Entrepreneurship education has several features that make assessing both professional and educational impacts particularly difficult. Most prominently, entrepreneurship is an outcome that can occur at varying times during a career. While there are examples of individuals pursuing entrepreneurship in college or immediately after graduation, research shows that it more commonly occurs when individuals reach their early 40s and have obtained the professional experience necessary to identify and pursue entrepreneurial opportunities (Azoulay et al., 2018). From a methodological perspective, the ability to establish causal relationships between an educational intervention and a professional outcome is difficult given the large number of factors influencing career choice, as well as challenges associated with conducting rigorous experimental research (Bauer, 2004). This is why the evaluation of entrepreneurship education, particularly in higher education, rests largely on self-reported measures (e.g., entrepreneurial intention or self-efficacy) rather than more objective ones (e.g., number of startup firms).

The expansion of entrepreneurship education programs to academic disciplines beyond business is another feature that adds complexity to program evaluation from both a methodological and practical perspective. Today, courses and programs are offered to a far more heterogeneous population of undergraduate and graduate students than ever before. They represent many academic units, demographic backgrounds, and participants with widely varying levels of business literacy and urgency to start a business (Morris et al., 2013). This expansion has led to its delivery in many formats, including full-semester courses, multi-course programs, noncredit workshops, and activities targeting particular topics or specific audiences, encompassing both traditional and experiential pedagogical approaches. Consequently, participants have different interests in and motivations for studying entrepreneurship beyond creating startups (Duval-Couetil et al., 2014; SÃ¡ & Holt, 2019). While some have venture ideas they wish to pursue in the short term, many others believe it will benefit them professionally at some point in the future. This leads to heterogeneity in disciplinary, curricular, and pedagogical approaches making it more difficult to compare effectiveness and outcomes across populations and settings. As a result, many educational stakeholders rely on intuition and assumptions informed by surveys, course evaluations, and qualitative student feedback to validate educational and career outcomes, rather than rigorous educational evaluation methods and study designs.

Descriptive Validity and Impact Evaluation

Validity is the extent to which a measurement, concept, or conclusion accurately reflects the real world. In impact evaluation, validity is an important consideration because data are collected via

	Key Question	Examples of Violations
Internal validity	Does the intervention really cause a change in the outcome?	• Preexisting differences between treatment and control conditions • Some event influencing the effect at the same time as the intervention • Pre-test measure influencing the effect • Differential loss of units
External validity	Is a causal relation generalized across different settings?	• Limited causal relations, only effective to some types of people and situations
Construct validity	Are the operational definition and measurement of the theoretical constructs appropriate?	• Weak evidence of validity and reliability of outcome measures • Contamination of the treatment
Statistical conclusion validity	Are the conclusions about causal relations based on appropriate statistical results?	• Insufficient statistical power (small sample sizes) • Inappropriate statistical techniques • Use of redundant statistical tests ("fishing expedition") • No reporting effect sizes and confidence intervals
Descriptive validity	Are key features of the evaluation study presented in a research report?	• A lack of information in evaluation reports (e.g., sample size, attrition rates, research design, description of treatment)

Figure 1. Validity criteria for evaluation research quality and examples of violations.

operationalized conceptual frameworks, instruments, and surveys that are designed to estimate the intended outcomes of a particular educational intervention, and counterfactual outcomes (i.e., outcomes in the absence of the intervention). Because unobservable and unexpected factors are potentially involved in experimental or quasi-experimental designs, securing evidence of validity is a requisite for causal inference, but is also a serious challenge in impact evaluation (Rossi et al., 2019). Cook and Campbell (1979) developed a foundation for understanding and controlling substantive and methodological biases in evaluation research through a better definition of validity, and a greater understanding of factors that bias data collection, analysis, and interpretation in social science. They proposed four criteria for evaluating methodological quality, including statistical conclusion validity, internal validity, construct validity, and external validity. These serve as a foundation for designing and evaluating evaluation research in a way that identifies and prevents potential validity threats, thereby allowing us to make valid causal inferences through the evaluation research (Figure 1).

Descriptive validity, a fifth criterion, was proposed later (Farrington, 2003; Lösel & Köferl, 1989). Descriptive validity refers to the adequacy of reporting critical elements of a research design, and the manner in which evaluation researchers secured all sources of validity evidence for a particular study. This involves reporting and describing study details with transparency, specifying the strengths of the research and describing potential threats to validity in terms of overall research design, sample, measures, data, and data analysis. Given variations in methodological rigor, this information is essential for reviewers, policymakers, and other research consumers wishing to judge the quality and generalizability of an individual study. Descriptive validity is essential when an impact evaluation study is disseminated broadly (e.g., journal

publication) or included in research syntheses such as meta-analysis and systematic review research (Gill, 2011; Rossi et al., 2019).

Researchers have recommended minimum elements that should be reported to assess the quality of evaluation research. These include the characteristics of experimental units and settings; sample sizes; hypotheses tested; descriptions of interventions; program implementation and delivery; descriptions of treatments and control conditions; operational definitions and measurement of outcomes; reliability and validity of outcome measures; follow-up period(s) after an intervention; statistical methods and effect sizes; independent and extraneous variables; and potential conflicts of interest (Boruch, 1997; Rossi et al., 2019). These elements alone do not ensure the quality of evaluation research; rather, they represent the minimum necessary to approximate conventional scientific research (Figure 1).

Some academic disciplines and organizations have embraced standards as a way to alleviate issues arising from inadequate reporting of research results. In the health and medical sciences, the Consolidated Standards of Reporting Trials (CONSORT) created a checklist of items to be included when reporting the results of a randomized trial. This checklist was also adopted by the American Psychological Association and many leading journals in the health sciences (Moher et al., 2012; Schulz, Altman, Moher, 2010). It even inspired the field of criminology to apply them to a range of criminal justice interventions (Gill, 2011). For educational programs, the National Science Foundation (NSF) in the United States provides guidelines for evaluation reporting. These highlight the importance of including background information (e.g., problems addressed, location, resources used to implement the project), evaluation study questions, evaluation procedures (e.g., sample selection procedures, representativeness of the sample, use of comparison or control groups), findings, and conclusions with recommendations (Frechtling et al.).

In entrepreneurship education, it is often impossible to design and implement evaluation studies with the elements of rigorous experimental research design, considered the "gold standard" (Henry, 2015). Therefore, descriptive validity is especially important for entrepreneurship education stakeholders and research consumers wishing to determine the inferential power of individual evaluation studies. Despite its potential value, descriptive validity has not been discussed in the entrepreneurship education literature, where the availability of guidelines and standards would likely improve the overall quality of evaluation research, as they have in other fields (Moher et al., 2012; Plint et al., 2006).

Impact Evaluation Research in Entrepreneurship Education

"Does a new or existing program result in a change?" is a key question and motivation for conducting impact evaluation research (Royse et al., 2016). Although evaluation research varies greatly in its scope, approaches, and methods, its primary purpose is to provide objective, reliable, and systematic evidence that supports whether a program successfully attains its stated goals. We are not the first scholars to discuss the rigor of evaluation research in entrepreneurship education. Instead, we are building on significant and important prior work. Recent systematic review and meta-analysis studies provide valuable insights to identify conceptual, methodological, and experimental design issues across research studies (e.g., Bae et al., 2014; Grégoire et al., 2019; Martin et al., 2013; Nabi et al., 2017; Rideout & Gray, 2013; for review, see Appendix A).

Systematic review and meta-analysis methods are research synthesis techniques, the goal of which are to "present the state of knowledge concerning the relation(s) of interest and to highlight important issues that research has left unsolved" (Cooper et al., 2009, p. 4). Systematic review is a qualitative approach that reduces bias by identifying, appraising, and synthesizing studies on a particular topic with the goal of synthesizing issues, prevalent practices, and future

directions for research (Uman, 2011). They can include a meta-analysis that uses statistical techniques to synthesize data from several studies into a single quantitative estimate to understand trends across studies. Nabi et al. (2017), Rideout and Gray (2013), and Grégoire et al. (2019) conducted systematic reviews of published studies in entrepreneurship education, where they synthesized, reviewed, and critiqued evaluation research and experimental design. Bae et al. (2014) and Martin et al. (2013) used meta-analysis. Below, we discuss the perspectives taken by these researchers and the methodological issues identified in their studies.

Rideout and Gray (2013) conducted a methodological critique of empirical research in entrepreneurship education to determine whether it really "works." The authors used Storey's model (2000) for evaluation validity, which characterizes descriptions of participants and their opinions about a program as "monitoring," not "evaluation." The authors found that only 12 studies met their criteria for robust impact evaluation, concluding that more research was necessary to determine confidently whether entrepreneurship education does indeed "work." The authors had several suggestions for making evaluation research inferentially meaningful. Most notably, these included providing comparisons between participants and nonparticipants and involving matched control groups to minimize self-selection bias. Other recommendations included employing more powerful quasi-experiential and experimental designs, longitudinal designs, larger sample sizes, psychometrically validated measures, and consideration of other validity threats.

Nabi et al. (2017) reviewed 159 published papers to examine the relationship between pedagogical approaches and outcomes in entrepreneurship education. They used a theoretical framework based on impact indicators (Jack & Anderson, 1998) to evaluate each study. According to the framework, any temporary, subjective, or personal changes measured during a program were considered lower-level impact indicators, and any long-term objective and socioeconomic impacts such as startups, venture performance, and contributions to the economy and society were considered higher-level indicators. The authors found that most research used short-term and subjective outcome measures, with weak links to pedagogical approaches. They noted that long-term impact measures were scarce in entrepreneurship evaluation research.

Most recently, Grégoire et al. (2019) evaluated 144 individual studies with regard to validity, robustness, and relevance of experimental design. Their scope was broader than education evaluation, but they provided various insights to improve the quality of evaluation research through rigorous experimental designs. They highlighted methodological challenges, including weak designs, analytical techniques, and the use of measures without evidence of construct validity. The authors also noted a lack of ecological and external validity, meaning that the methods, materials, and settings were limited in their ability to approximate real-world settings, or be generalized to other situations, people, stimuli, and times (Mitchell & Jolley, 2001).

In a meta-analysis study, Martin et al. (2013) investigated the outcomes of entrepreneurship education and training programs through the lens of human capital theory. Given conflicting findings related to impact in the literature, they sought to examine it quantitatively by analyzing 42 individual studies considered rigorous because they employed both pre- and post-measures, as well as treatment and control groups (Cook & Campbell, 1979). In their analysis, outcome variables were classified into two categories: (a) human capital assets, which included more indirect measures such as knowledge and skills, competency, motivation, and entrepreneurial intentions; and (b) entrepreneurship outcomes, which included more direct and actual measures of entrepreneurial behaviors and activities such as startup activity and financial success. The authors found that, overall, entrepreneurship education is positively related to both. They also found that positive outcomes were more robust for academic-focused interventions than for training-focused interventions. They indicated that many evaluation studies were not included in their review because of methodological and reporting issues, and they felt that low-quality studies might lead to an overestimation of entrepreneurship education's impact. In particular, they

addressed the important role of moderating variables (e.g., age, gender, ethnicity, type of educational intervention) in evaluation research, which can cause differential effects on overall impact and, if controlled adequately, help establish the relationships more clearly and accurately.

Bae et al. (2014) combined the effect sizes of 73 empirical studies to examine the effect of entrepreneurship education on entrepreneurial intention. The authors highlighted the issue of self-selection bias, meaning that greater entrepreneurial intention could be the result of a pretendency among students who study entrepreneurship to want to be entrepreneurs. They found that the significant effect that entrepreneurship education had on entrepreneurial intentions, indeed, depended on the process of sorting students by their predisposition for entrepreneurship education, thereby making a determination of causation more difficult.

Our review of evaluation standards and validity criteria identified key information necessary to appraise the quality of experimental and quasi-experimental designs in impact evaluation research. Prior synthesis studies in entrepreneurship highlighted the significant variations in reporting practices, and concerns about the methodological rigor of impact studies (Bae et al., 2014; Grégoire et al., 2019; Martin et al., 2013; Rideout & Gray, 2013). Our review also indicates that validity, and in particular descriptive validity, has been overlooked in prior research synthesis studies. In light of this, our goal is to arrive at standards and best practices for impact evaluation in entrepreneurship education research so that educators, scholars, professional associations, journal editors, and funding agencies have a checklist of items that should be included in studies and reports that evaluate outcomes and impact (Farrington, 2003).

A Framework for Evaluating Descriptive Validity

Inspired by research guidelines and descriptive validity elements (Boruch, 1997; Moher et al., 2012; Schulz, Altman, Moher, et al., 2010), we developed a list of key criteria for evaluating impact research in entrepreneurship education that we refer to as the descriptive validity framework (Figure 2). Our framework includes six categories of key methodological elements pertinent to designing and reporting evaluation research, including research design, sampling and grouping, intervention, outcome measures, statistical methods and results, and confounding variables. We integrated the methodological issues identified in prior entrepreneurship syntheses studies into the framework. In Figure 2, we explain the importance of each category to evaluation research in detail.

Research Design

Impact evaluation involves measuring the effect of an intervention by utilizing experimental or quasi-experimental designs. Educational program evaluation can be viewed as a structured comparison designed to examine relevance, effectiveness, and impact in light of specific objectives. To investigate causal relations, it is necessary to incorporate counterfactual outcomes in evaluation research. This is usually accomplished by measuring changes in outcome variables (a) over time (repeated measures) and (b) in comparison to one or more comparison or control groups. However, prior research syntheses found that many studies did not employ such comparisons. Instead of program effects, they provided baseline characteristics of participants (Martin et al., 2013; Nabi et al., 2017; Rideout & Gray, 2013). A limitation of this approach is that it restricts evaluation research to merely "monitoring" the progress of program participants, rather than providing inferential evidence related to the program impact (Rideout & Gray, 2013; Royse et al., 2016; Storey, 2000).

Therefore, evaluation research should articulate how a study is designed to measure differences in outcomes caused by a specific educational intervention. Either a repeated measures

Category	Key Elements for Reporting in Evaluation Research	Methodological Issues in Entrepreneurship Research
Research design	○ Descriptions of the overall research design ○ Hypotheses based on theoretical foundations	• A lack of experimental or repeated/longitudinal design [G,M,R]
Sampling/grouping	○ Sample selection methods ○ Description of group assignment ○ Baseline demographic information of both groups ○ Baseline equivalency between the groups ○ The sample size for each group ○ Attrition rates	• Experiments with samples that do not represent real-life settings or target populations [G] • Small sample sizes [R] • Comparisons between treatment and control groups with no consideration of self-selection bias or matching groups (self-selection bias) [B,M,R]
Intervention	○ Description of treatment and control conditions received (e.g., duration, goals, content, deliverers) ○ Characteristics of experimental units and settings	• Intervention not aligned with pedagogy [N] • Heterogeneity of the manipulation/intervention within the same group [B,G]
Outcome measures	○ Evidence of validity for outcome measures ○ The rationale for measuring the outcomes aligned with the goals of the intervention	• Weak evidence of construct validity for outcome measures (constructive validity of measurement) [G,N,R] • Too much relying on indirect and psychological outcomes only (not measuring direct outcomes) [N,R]
Analytical methods and results	○ Description of statistical methods used to examine the group/time differences ○ Reports of statistical results for precisions (e.g., effect sizes, confidential intervals)	• No reporting or discussion of effect sizes or the magnitude of observed relations [G]
Confounding variables	○ Theoretical foundations for possible extraneous variables impacting the causal relations ○ Examination of confounding/control variables	• No control for within-group bias [B,G]

Figure 2. Descriptive validity framework for entrepreneurship education research.

Note. Methodological issues were identified in the following research synthesis studies: B = Bae et al. (2014); G = Grégoire et al. (2019); M = Martin et al. (2013); N = Nabi et al. (2017); R = Rideout and Gray (2013).

design or a group comparisons design is a bare minimum for impact evaluation. Gill (2011) recommended that the research design be clearly stated in the abstract of a published evaluation paper. Clear specification and reporting of research objectives, preferably including hypotheses and evaluation questions, are also recommended so that research consumers can determine what an individual study attempted to achieve (Gill, 2011; Moher et al., 2012).

Sampling and Grouping

Sampling is critical to ensure the generalizability of research findings and make inferences from comparisons with counterfactual conditions. One criticism of entrepreneurship education evaluation research is that it often involves small sample sizes or specific groups of participants, and consequently does not have sufficient statistical power (Rideout & Gray, 2013). These weaknesses and sample selection biases are some of the unique aspects of entrepreneurship education, which tends to target groups of people and circumstances that do not necessarily reflect real-life populations or settings (Grégoire et al., 2019). Matching group techniques or efforts to ensure group equivalency also should be considered before interventions (Bae et al., 2014; Martin et al., 2013; Rideout & Gray, 2013). Therefore, reliable evaluation research should employ and report sample selection methods, group assignment methods, equivalency between treatment and comparison groups, sample sizes for both groups, and attrition rates.

Intervention

In evaluating educational programs, the intervention is the educational program itself. It serves as an independent variable, which is hypothesized to be the leading cause of changes in outcome measures. However, there are variations in how individual studies define and measure interventions. In some cases, studies rely on participants' self-rated experiences, where the independent variable is heterogeneous among participants. For example, if a researcher designates the

independent variable to be *an educational experience in entrepreneurship education*, the intervention could be a composite of a degree program, a course, or even a 3-day workshop. In this case, it is unclear what we can infer about the impact of one specific program.

Bae et al. (2014) brought up the issue of *common method variance* and *common rater effect* when using self-reported measures, particularly for designating and measuring independent variables. *Common method variance* is the "variance that is attributable to the measurement method rather than to the constructs the measures represent" (Podsakoff et al., 2003, p. 879), and *common rater effect* is the inflated correlation between an intervention and outcomes when the subjects of the two measures are the same. Thus, if both the independent and dependent variables are self-rated in evaluation research, the results are likely to be inflated. Therefore, impact evaluation research should clearly describe the treatment and control interventions that participants received, including duration, goals, contents, and deliverers of the intervention, as well as how researchers designated and measured the independent variable.

Outcome Measures

The impact of an educational program is measured as the difference between outcome levels for an individual exposed to a particular intervention, and one exposed to a counterfactual condition. In entrepreneurship education, outcome variables are often measured with self-reported, indirect, or subjective measures (e.g., entrepreneurial intention, self-efficacy) rather than more direct or objective measures (e.g., number of startup firms; Nabi et al., 2017; Rideout & Gray, 2013). Rossi et al. (2019) categorized the outcomes by their attainability. Proximal outcomes are expected to be attained immediately and directly from a program, but they are often not the ultimate outcomes that the program intends to make. Distal outcomes are direct outcomes, but they are difficult to measure. Ideally, measuring a balance of both proximal and distal outcomes would be beneficial and interpretable in evaluation research (Rossi et al., 2019).

An outcome measure with weak validity evidence is also a major concern pointed out in prior studies (Grégoire et al., 2019; Nabi et al., 2017; Rideout & Gray, 2013). Without information about the quality of outcome measures, consumers of evaluation research cannot be assured of the findings. Therefore, a detailed description of outcome measures, validity evidence including reliability coefficients, results of factor analysis, and efforts to improve the outcome measures should be reported in an evaluation study.

Statistical Methods and Results

In impact evaluation research, statistical analysis is used to test differences in outcome measures. Prior research synthesis studies discussed the need for more robust and accurate analytical modeling in entrepreneurship education. They also pointed out that individual studies did not report and discuss effect sizes or magnitude of observed relations (Rideout & Gray, 2013), which enable us to arrive at more precise estimations of impact. Statistical significance, effect sizes, confidence intervals, as well as the rationale for choosing particular statistical methods, should be reported in evaluation research.

Confounding Variables

Confounding refers to a relationship between an intervention and outcome that is falsely obscured or accentuated by a third variable, which often is a characteristic of participants (Meinert, 1986). In impact evaluation research, the differences in outcome measures between treatment and comparison groups can systematically differ due to confounding variables. Thus, it is important to

account for the systematic differences in baseline characteristics when estimating the effect of an intervention. Regression adjustment is often used in applied research to account for the effect of confounding variables, which significantly influence the effect of the treatment (MacKinnon et al., 2000). Bae et al. (2014) argued that research in entrepreneurship education often did not control for within-group bias, which can be addressed by utilizing confounding variables. Therefore, it is recommended that evaluation research include the theoretical foundation or reasoning for the impact extraneous variables could have on causal relations, and descriptions of confounding and control variables.

Methods

Using the descriptive validity framework and criteria in Figure 2, we examined impact evaluation studies in entrepreneurship education to identify the methodological practices that are most prevalent in the field. Below, we describe our methods.

Coding Scheme

Unlike research synthesis studies that examined whether entrepreneurship education was effective, our primary goal was to identify prevalent methodological and reporting practices with respect to how impact was measured. Table 1 presents the coding scheme we developed, based on requirements for descriptive validity that we used to examine individual studies.

Search Strategy

We conducted a systematic review of impact evaluation research studies in entrepreneurship education to identify peer-reviewed journal articles published between 1998 and 2019. We operationally defined *impact evaluation research* as empirical studies attempting to measure changes in outcomes as evidence of the impact of entrepreneurship education. Our focus was university-based programs to minimize variability in program characteristics and target populations. We examined quantitative approaches that clearly entailed measuring or assessing the impacts, effectiveness, or outcomes of a program. The criteria for inclusion in our review were as follows:

1. Empirical studies that measured the impact of a university-based entrepreneurship education program (e.g., college course, training program, degree program, workshop), where the program served as an intervention or a condition of a treatment group;
2. Studies where the focus and purpose were to measure the impact of a program, rather than report on survey/instrument development, or program overviews, where assessment or characteristics of program participants may be acknowledged;
3. Studies that quantitatively examined at least one outcome measure, which was hypothesized to change, increase, or decrease, as a result of an entrepreneurship education program;
4. Completed studies that were published in the form of peer-reviewed journal articles written in English.

Using the keywords "entrepreneurship education," "impact," "outcomes," "measures," "university," and "evaluation," we conducted a database search to identify journal articles (Science Direct, EBSCO, Google Scholar). We also searched among the leading peer-reviewed entrepreneurship journals identified by the Association to Advance Collegiate Schools of Business (AACSB), which included *Academy of Management Learning and Education, Entrepreneurship*

Table 1. Coding Scheme.

Category	Key elements	Coding
General information		• Author/year/journal
Criteria for inclusion		• Description of the study purpose
Research design	Details about research design	• Overall design (e.g., pre- and post-design, quasi-experimental design) • Number of repeated measures • Duration between repeated measures • Use of a control group • Follow up
	Hypotheses based on theoretical foundations	• Specification of a hypothesis
Sampling/grouping	Sample selection methods	• Specification of methods reducing self-selection bias (e.g., randomized design)
	Sample characteristics	• Sample size • Ages • Majors • Nation where the study conducted
	Equivalence between treatment and control groups	• Definition of a control group
Intervention	Details about intervention	• Description of intervention • Duration of the program
	Homogeneity within a group	• Description of condition applied to each group
Outcome measures	Alignment with goals	• Description of outcome variables
	Evidence of measurement validity	• Number of questions • Specification of validity evidence (e.g., reliability, relations to other variables)
Statistical methods and results	Rigorous method aligned with the purpose	• Specification of statistical methods
Confounding variables	Examination of confounding/control variables	• Control/extraneous variables • Interaction effects

Theory and Practice, International Entrepreneurship Management Journal, Journal of Business Venturing, Journal of Entrepreneurship, Journal of Entrepreneurship Education, Journal of International Entrepreneurship, Journal of Small Business Management, and *Technovation*. At first, we searched for studies with abstracts stating their purpose was to evaluate, investigate, or examine the impact of an entrepreneurship education program, which yielded 207 studies. However, many studies were excluded from our review because they did not use quantitative approaches (e.g., used descriptive or qualitative approaches instead) or did not meet at least one of our criteria (e.g., did not describe the characteristics of program participants, did not focus on a specific intervention, or took place in secondary schools). Our search resulted in 61 empirical studies that met our criteria. This was not surprising given that prior review studies noted weak study designs in entrepreneurship evaluation research (e.g., Martin et al., 2013; Rideout & Gray, 2013). Table 2 summarizes the studies included in our review.

Table 2. Summary of Studies Included in the Review.

First author	Year	Program duration	Repeated	Comparison Group	Randomized assign	Sample size	Country
Repeated × Group Comparison							
DeTienne	2004	13 weeks	Y	Y	Y	130	USA
Gielnik	2015	12 months	Y	Y	Y	384	Uganda
Gielnik	2017	12 weeks	Y	Y	Y	227	Kenya
Alharbi	2018	1 semester	Y	Y	N	523	Saudi Arabia
Hansemark	1998	9 months	Y	Y	N	70	USA
Jahani	2018	3 sessions	Y	Y	N	76	Iran
Karlsson	2013	1.5 months	Y	Y	N	103	Denmark
Ohland	2004	2 semesters	Y	Y	N	91	USA
Oosterbeek	2010	1 year	Y	Y	N	562	Netherlands
Rauch	2015	52 weeks	Y	Y	N	142	Netherlands
Sanchez	2011	8 months	Y	Y	N	863	Spain
Souitaris	2007	5 months	Y	Y	N	250	UK, France
Thursby	2009	2 years	Y	Y	N	217	USA
No Repeated × Group Comparison							
Akin	2015	1 semester	N	Y	N	920	Turkey
Asghar	2019	—	N	Y	N	542	Finland
Bhat	2018	—	N	Y	N	350	India
Solesvik	2014	—	N	Y	N	243	Ukraine
Premand	2016	20 days	N	Y	Y	1580	Tunisia
Westhead	2016	—	N	Y	Y	175	Ukraine
Repeated × No Group Comparison							
Karimi	2016	4 months	Y	N	—	205	Iran
Alcaraz-Rodriguez	2014	15 weeks	Y	N	—	17	Mexico
Boukamcha	2015	20 days	Y	N	—	240	Tunisia

(Continued)

Table 2. Continued

First author	Year	Program duration	Repeated	Comparison Group	Randomized assign	Sample size	Country
Chang	2013	12 weeks	Y	N	—	27	UK
Fayolle	2006	1 semester	Y	N	—	20	France
Fayolle	2015	3 days	Y	N	—	158	France
Fretschner	2013	1 semester	Y	N	—	62	Germany
Hattab	2014	14 weeks	Y	N	—	182	Egypt
Rossano	2016	1 semester	Y	N	—	150	Germany
Shealy	2018	1 semester	Y	N	—	31	USA
Shinnar	2014	2 semesters	Y	N	—	317	USA
Stamboulis	2014	2 semesters	Y	N	—	169	Greece
Von Graevenitz	2010	1 semester	Y	N	—	343	Germany
Vorley	2016	14 days	Y	N	—	60	UK
			No Repeated × No Group Comparison				
Dutta	2011	—	N	N	—	221	USA
Elmuti	2012	—	N	N	—	170	USA
Farashah	2013	1 semester	N	N	—	628	Iran
Gerba	2012	—	N	N	—	156	Ethiopia
Harms	2015	—	N	N	—	172	Netherlands
Kiyani	2017	1 semester	N	N	—	100	Pakistan
Lanero	2011	—	N	N	—	800	Spain
Lee	2005	—	N	N	—	377	US, Korea
Maresch	2016	—	N	N	—	3581	Austria
Passoni	2017	—	N	N	—	491	Brazil
Pruett	2012	3 sessions	N	N	—	105	USA
Solesvik	2013	—	N	N	—	189	Europe
Storen	2014	—	N	N	—	2827	Norway

(Continued)

Table 2. Continued

First author	Year	Program duration	Repeated	Comparison Group	Randomized assign	Sample size	Country
Walter	2013	—	N	N	—	1530	Germany
Zainuddin	2010	—	N	N	—	186	Malaysia
Otache	2019	—	N	N	—	256	Nigeria
Mamun	2017	—	N	N	—	375	Malaysia
Arranz	2017	—	N	N	—	1475	Spain
Baidi	2018	—	N	N	—	500	Indonesia
Kirby	2013	—	N	N	—	60	Egypt
Lourenco	2011	4 sessions	N	N	—	384	UK
Zhang	2014	—	N	N	—	494	China
Hasan	2017	—	N	N	—	200	Bangladesh
Varamäki	2015	—	N	N	—	197	Finland
Lefebvre	2013	—	N	N	—	50	France
Saji	2018	—	N	N	—	63	UAE
Sun	2016	—	N	N	—	201	Hongkong

Category	Key Elements	Examples of Practice		Potential Sources for Validity
		Weak	Stronger	
Overall research design	Details about research design	Treatment group only One-time measure	Treatment and control groups, longitudinal measures	Internal
	Hypotheses based on theoretical backgrounds	Lack of hypothesis	Clear hypotheses based on theoretical backgrounds	
Sampling/grouping	Sample selection methods	Convenience sampling Snowball sampling	Randomized/stratified sampling	External, Internal
	Sample characteristics	Less information (demographic only)	Evidence of population representativeness	Internal, Descriptive
	Equivalency between treatment and control groups	Lack of evidence	Randomized controlled trials, matched grouping	Internal, Statistical
Intervention	Details about intervention	General experience in entrepreneurship education	Specification of purpose, components, targets, delivery methods, and delivery personnel	Descriptive, Internal
	Homogeneity within treatment/control groups	Ignoring the heterogeneity of interventions/experiences in a treatment group	Information about the homogeneity of a group	Statistical, Internal
Outcome measures	Alignment with goals	Indirect and short-term outcomes without alignment of intervention (e.g., perceived intention, motivation)	Intended direct and long-term outcomes (e.g., actual start-ups)	Statistical, Descriptive
	Evidence of measurement validity	Lack of evidence	Multiple sources of evidence	Construct
Analytic methods and results	Rigorous method aligned with the purpose	Descriptive statistics T-test, ANOVA, repeated t-test	Structural equation modeling, Multilevel modeling, Longitudinal modeling	Statistical, Descriptive
	Information about the statistical accuracy	Lack of information	Multiple information (effect sizes, confidential intervals)	
Confounding variables	Examination of confounding/control variables	Lack of confounding variables	Examination of multiple confounding variables based on theoretical backgrounds	Internal

Figure 3. Impact Evaluation Research Standards (source: authors) .

Results

This section summarizes the most prevalent methodological practices and issues found in the 61 studies we reviewed. Our results are organized into the six categories from our descriptive validity framework (Figure 2) and are synthesized in a set of Impact Evaluation Research Standards (Figure 3).

Research Design

An essential component of evaluation research is a comparison either across time or between groups. Good practice entails providing a clear description of how a study employed a comparison to address the research question(s). A statement of research hypotheses, developed via a thorough literature review, is also strongly recommended. Our review showed that most studies (78.7%) stated research hypotheses before presenting their research methods and findings. A preferable practice is to describe the research design (e.g., randomized controlled experimental design, quasi-experimental design) in the abstract of the study (Gill, 2011). This avoids confusion for potential readers and evaluation consumers, who will better understand from the start, the causal relations the authors hypothesized and attempted to examine. For example, DeTienne and Chandler (2004) stated in their abstract that they employed "a variation of a Solomon Four Group Designed experiment." This helps readers search for and navigate relevant information throughout the paper and better understand the research framework. However, in 21.3% of the studies we reviewed, the research design was not clearly stated, and in some cases, there were disparities between the purpose specified in the abstract and the actual research design the study used.

Table 3. Overall Research Design Implemented by the Studies.

Research design	Before 2012 (*N* = 16)	2012–2019 (*N* = 45)	Total (*N* = 61)
No comparisons	5	14	19 (31%)
Repeated measure only (2 times measures)	2	11	13 (21%)
Longitudinal but no comparison group (+3 times measures)	0	3	3 (5%)
Comparison group only	1	11	12 (20%)
Comparison group × repeated measure (2 times measures)	7	4	11 (18%)
Comparison group × longitudinal (+3 times measures)	1	2	3 (5%)

Table 3 summarizes the research designs used in the 61 studies in our sample. Although it has been two decades since Storey (2000) critiqued entrepreneurship education research for not employing appropriate comparisons, almost a third of the studies we reviewed still did not measure substantial changes resulting from an intervention, over time, or with a comparison group. Interestingly, more recent studies fared no better; only 13.3% of the studies published after 2012 (six out of 45) used relatively rigorous research designs, as compared to 50% (eight out of 16) for those published before 2012. One possible reason is the use of advanced statistical techniques, since most studies without a comparison group used structural equation modeling (SEM; e.g., Boukamcha, 2015; Karimi et al., 2016; Lourenço & Jayawarna, 2011; Shinnar et al., 2014). These studies attempted to examine the relationship between program characteristics, or the number of students taking entrepreneurship education courses, to outcome variables. However, without substantial comparisons across time or groups, these studies could not tell us what changes the programs brought to program participants. Therefore, although SEM is a sophisticated statistical method that controls measurement error and is statistically powerful with large sample sizes, it does not compensate for fundamental limitations in research design.

Sampling and Grouping

A clear description of sampling and grouping methods enables potential readers to evaluate the generalizability of an impact study. Thus, individual studies should report sample sizes for each group, attrition rates if the study used repeated measures, and relevant baseline information for the samples. All 61 studies we reviewed reported their total sample sizes; however, a few studies did not clearly report the number of participants in the treatment and comparison groups separately (e.g., Alharbi et al., 2018; Thursby et al., 2009a, 2009b). In addition, where studies use repeated measures, sample sizes for each time point measured (e.g., pretest, post-test), and attrition rates for each group should be reported. Even in a semester-long program, attrition is inevitable due to students dropping out, absences on the date of post-measures, or a lack of availability for follow-up. Among the 21 studies that used repeated measures, 11 studies did not report the number of participants for each time period.

The studies we reviewed represented 25,217 participants in total. Not surprisingly, the variance in sample sizes across the studies was considerable, ranging from 17 to 3581, with an average of 420. Twelve studies had sample sizes of less than 100, which may result in weakened statistical power. To generalize the results of evaluation research, it is best to collect data from a sample closely reflecting the characteristics of the target population. However, most studies (80.3%, $n = 49$) relied on availability or convenience sampling. A few of the more recent studies used large-scale data collected by randomized or stratified sampling methods (e.g., Gielnik et al., 2015; Hasan et al., 2017; Lanero et al., 2011; Westhead & Solesvik, 2016; Walter et al., 2013).

Table 4. Measures Associated with Interventions (*N* = 61).

Definition of interventions	Measures	*N* (%)
Participation in a designated intervention	Degree or long-term program (≥2 semesters)	17 (28%)
	A course (1 semester)	16 (26%)
	Short-term program (<1 semester)	6 (10%)
Experience in entrepreneurship education	Binary (yes, no)	16 (26%)
	Number of experiences	3 (5%)
	Rating perceived experiences (Likert-scale)	3 (5%)

For example, Hasan et al. (2017) used simple random sampling with 200 students, and Lanero et al. (2011) used stratified sampling to secure the equivalency of the samples to the populations. In four other studies (DeTienne & Chandler, 2004; Gielnik et al., 2015, 2017; Premand et al., 2016), the authors randomly assigned participants to comparison groups, or randomly selected a control group to match with a treatment group.

Intervention

To measure the impact of an entrepreneurship education program, researchers should define a specific educational intervention as the independent variable. In most of the studies we reviewed ($n = 39$, 64%), the intervention was simply defined as participation in a specific type of entrepreneurship education program, as indicated in Table 4 (e.g., college course, training program, degree program, workshop). The benefit of this approach is that it controls the variance across the different program formats and types of interventions that participants received. As such, all members in a group share a common experience, in what is considered an equivalent program, that serves as the intervention for the study. We sorted entrepreneurship education programs into three categories depending on their duration: long-term programs (two or more semesters), one semester programs, and short-term programs (e.g., workshops). At the very least, the studies included in our review provided a brief description of the intervention, including program duration, goals, objectives, and activities. It is important to highlight the importance of reporting program characteristics as variance across interventions is expected.

Another approach used by researchers is to define the independent variable as participants' *experience* in entrepreneurship education, more generally, without regard to a specific program's characteristics. This method was used in studies measuring the impact of entrepreneurship education with secondary data, not directly collected from an experimental evaluation study, or in studies with a sample of heterogeneous groups of participants. Sixteen studies arrived at the independent variable using a binary or categorical scale by asking whether the participants had experienced one or more entrepreneurship education programs (e.g., "0" = no experience; "1" = one or more experiences). Since these studies relied on participants' subjective responses to define the intervention (i.e., what constitutes an experience), variance inevitably increased within the independent variable of the study. In other words, the particular characteristics and properties of a specific entrepreneurship education program (e.g., pedagogical approaches, program elements, the length, and required commitment to the programs) could differ widely among individuals in the same group. To control for within-group bias, participants should be asked about specific factors that might cause variance (i.e., the specific properties of entrepreneurship

education programs in which they participated) so that these can be controlled for in the analysis to measure the effect more accurately.

Outcome Measures

Impact evaluation research involves selecting outcome measures to represent changes resulting from exposure to a program. Therefore, the relevance and accuracy of these measures should be presented in evaluation reports. The relevance of different kinds of outcome measures has been discussed in the entrepreneurship education literature (e.g., Nabi et al., 2017; Rideout & Gray, 2013), and our review showed that most outcome variables were self-reported subjective or indirect measures rather than more objective measures (e.g., founding a startup). These fell into four categories: attitudes and motivations, knowledge and skills, behavioral and action-oriented outcomes, and actual startup activities. Subjective measures included a variety of participants' perceptions related to acquiring entrepreneurial knowledge and skills, motivation for entrepreneurship, or for entrepreneurship education. Although these subjective measures are aligned with some stated goals of entrepreneurship education, behaviors and actions related to involvement in actual startup activities are more accurate measures as they reduce potential bias and variation in respondents' self-reported ratings. Only four studies that met the criteria for inclusion in our study, used objective measures.[2] Outcome variables in all 61 studies are summarized in Table 5. A total of 202 measures were identified, averaging 3.31 measures per study.

Attitudes and motivation were most often used to evaluate the impact of entrepreneurship education (Table 6). These refer to psychological variables representing internal changes that promote and support students' performance and intention for startup activities, including attitudes toward entrepreneurial behavior (Kolvereid, 1996; Liñán & Chen, 2009), risk-taking (Roman & Maxim, 2017), subjective norms (Kolvereid, 1996), entrepreneurial motivation (Souitaris et al., 2007), and perceived desirability (Chen & Tan, 2009). The appropriate use of subjective measures requires collecting integrative evidence to determine if the assessment was aligned with its proposed interpretation and utility (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014). However, the studies we reviewed provided limited information regarding the evidence of validity.

Knowledge and skills relate to the content delivered to students and are measured through changes in entrepreneurial self-efficacy, perceptions of opportunities, creativity, decision-making, and achievement either as an individual or team member. Self-reported forms of these measures provide information regarding how confident respondents perceive their abilities for particular tasks to be; however, they inherently have bias given that respondents may not accurately recognize their own abilities. Thus, like self-reported measures of behavior and action, it is uncertain to what extent these measures predict actual performance. For example, a recent study of engineering students by Blikstein et al. (2017) highlighted the importance of examining self-assessed confidence and actual performance as it related to engineering competencies. They found that students' actual performance was not correlated strongly with their self-assessed skills and competencies. This could be the case in entrepreneurship education as well, where a self-reported assessment of competencies might have limited predictive ability for actual performance. Future research should examine associations between self-reported measures and long-term and short-term objective measures of entrepreneurial competence.

None of the studies we reviewed reported any sources of construct validity evidence such as *content specification*, *expert judgment*, or *systematic observations of behaviors* (see the standards for construct validity, American Educational Research Association., et al., 2014). In most studies, reliability coefficients and the results of factor analysis served as evidence of a scale's

Table 5. Categories of Outcome Measures Used for Coding.

Category	Measure	Definition	Measure variables
Attitudes & motivation	Self-report	Psychological variables representing internal changes that promote and support students' performance and intention for startup activities	• Attitudes toward entrepreneurial behavior/self-employment • Perceived behavioral control • General entrepreneurial self-efficacy • Need for achievement • Perceived desirability • Perceived feasibility • Values • Locus of control • Motivation
Knowledge & skills	Self-report or Objective	Confidence, self-ratings or actual data about capabilities and capacity for particular tasks and knowledge	• Entrepreneurial skills • Know-how/what/who/why • Creativity • GPA in the course • Self-reported creativity • Self-reported leadership • Self-reported teamwork • Problem solving • Entrepreneurial competencies
Behavioral and action-oriented outcomes	Self-report	Behavioral tendencies related to entrepreneurial activities, the most common being entrepreneurial intention	• Entrepreneurial intention • Entrepreneurial action
Actual startup	Objective	Number of startup firms and actual indices of performance	• Business creation • Income

internal structure. In terms of validity evidence based on a *relation to other variables*, research has found that entrepreneurial intention is positively associated with entrepreneurial self-efficacy, proactiveness, and other motivational and psychological characteristics related to entrepreneurship (e.g., Fayolle & Gailly, 2015; Kolvereid & Moen, 1997; Souitaris et al., 2007; Sánchez, 2011b). However, there is a lack of research supporting the assumption that entrepreneurial intention indeed predicts actual entrepreneurial behavior (Bae et al., 2014; Douglas & Shepherd,

Table 6. Outcome Variables Used in Review Studies.

		Research design			
		No repeated	Repeated	Longitudinal	Total
Proximal	Attitudes & motivation	49	61	6	116
	Knowledge & skills	18	22	3	43
	Behavior & planning	20	11	4	35
Distal	Actual startup	4	2	2	8

Note. The numbers in the table indicate the number of outcome variables, not of studies. Most of the individual studies used multiple outcome measures.

Table 7. Analytical Methods Applied to Measure the Effect of the Intervention.

Analytic methods	Repeated measure	Group comparisons	*N*	Year of study
Descriptive statistics	–	–	1	2013
	–	C	1	2017
	R	–	2	2013, 2016
T-test/ANOVA/ MANOVA	–	–	3	2005, 2012, 2017 (group comparison but not control group)
	–	C	5	2004, 2015, 2016 (3)
	R	–	8	2006, 2010, 2014 (3), 2015, 2017, 2018
	R	C	5	2004, 2007 (2), 2010, 2011
Regression (GLS)	–	–	8	2010, 2011, 2012 (2), 2013, 2016, 2017, 2018
	R	–	1	2015
	–	C	3	2015, 2018 (2)
	R	C	3	2007, 2009, 2013
Nonlinear regression	–	–	3	2011, 2013, 2014
	–	C	1	2014
Hierarchical linear modeling	–	–	2	2013, 2015
	–	C	2	2011, 2014
Structural equation modeling	–	–	4	2010, 2011, 2016, 2019
	–	C	1	2019
	R	–	6	2013, 2014, 2015 (2), 2016, 2017
	R	C	1	2015
Growth curve modeling	R	C	1	2017

2002). Given that entrepreneurial intention is used frequently in evaluation studies, it is necessary to critically examine this hypothesis by conducting longitudinal research to determine whether a person whose entrepreneurial intention has increased during or immediately after an entrepreneurship education actually gets involved in startup activity in the future.

Analytical Methods

Table 7 summarizes how the 61 studies applied statistical methods to examine the effect of entrepreneurship programs. T-test/ANOVA ($n = 21$) and OLS regression analysis ($n = 15$) were used most frequently. However, the application of these methods provides limited information regarding the causal relations between treatment and control, and between pre- and post-treatment, particularly when inferentially weak research designs are used (e.g., single group pre- and post-test, heterogeneity between treatment and comparison groups; Rideout & Gray, 2013). For example, these methods do not consider measurement errors of latent variables, which are examined by confirmatory factor analysis, particularly for psychosocial outcome variables such as motivation, intention, and self-efficacy. Thus, if a latent factor model of an outcome measure emerges from confirmatory factor analysis, latent variable analysis (e.g., path analysis, SEM) is recommended as it uses measurement errors of latent variables for comparing

multiple groups rather than t-test or ANOVA (Aiken et al., 1994; Brown, 2015). Further, in cases where outcome variables are measured more than three times, and researchers are interested in examining systematic changes in the outcome variables due to the treatment over time, the use of longitudinal analytic techniques is recommended given that time is nested within a person (Singer et al., 2003).

More recently, studies have used hierarchical linear modeling (HLM) and SEM to measure the effect of entrepreneurship education in response to a need for more rigorous and accurate analytic modeling in evaluation research. Harms (2015) applied HLM to estimate the differential effects of entrepreneurship education with individual-level and group-level predictors and outcome variables (e.g., individual exam scores, group assessment). Gielnik et al. (2015) used a randomized controlled group design to examine the impact of entrepreneurship education training over 12 months and applied SEM to examine their hypotheses. Their analytical methods enabled them to simultaneously explore the association of multiple variables estimating the moderating, as well as direct effects, of predictors and outcome variables. However, we found no longitudinal studies based on HLM and SEM examining the growth of outcome variables over three or more periods of time, except for a growth curve modeling study (Gielnik et al., 2017), which is a method increasingly being used in educational research.

Controlling Confounding Variables

Among the 61 studies, 19 studies (31%) measured the impact of the entrepreneurship education intervention while controlling for the effects of confounding variables such as gender, age, entrepreneur family/relatives/friends, prior work/education experience, prior grades, geography, ethnicity, and religion. In terms of gender effects, female students were found to have lower exam scores (Harms, 2015; Zhang et al., 2014). Having family, relatives, or friends who are entrepreneurs was considered a confounding variable, but we found inconsistency in empirical research findings. In some studies, these variables had no significant confounding effects on entrepreneurial self-efficacy, entrepreneurial intention, or perceived competencies (Sánchez, 2011b, 2013; von Graevenitz et al., 2010). To summarize, because few studies identified and controlled for confounding variables in their analyses and there were inconsistencies in study results where demographic variables were concerned, we cannot be certain if these variables are confounders that lead to false results in evaluation studies.

Summary of Methodological Standards and Best Practices

Our review and analysis showed that many impact studies in entrepreneurship education lacked the essential methodological information that allows readers to gauge the quality of evaluation research. This led us to create a set of Impact Evaluation Research Standards to summarize guidelines and best practices for determining the methodological rigor of studies. We organized these standards into six categories: overall research design, sampling/grouping, intervention, outcome measures, analytical methods and results, and confounding variables (Figure 3). Each category presents examples of what are considered strong evaluation practices and approaches for demonstrating validity evidence, according to contemporary educational practices and standards (Farrington, 2003). We also propose best practices for more robust evaluation procedures through examples of reliable and rigorous research designs and techniques. Figure 3 can serve as a checklist for entrepreneurship educators, scholars, and research consumers wishing to improve education evaluation research.

Discussion

The purpose of this study was to identify and describe the specific methodological barriers limiting the ability of educators and scholars to demonstrate the impact of entrepreneurship education in a data-driven way. To do so, we synthesized the findings of prior studies in entrepreneurship in light of best practices in educational evaluation research. We created a framework comprised of standards for applying what are considered rigorous research methods to entrepreneurship education research. We then conducted a new systematic review, grounded in educational evaluation fundamentals, which allowed us to synthesize specific methodological practices and issues occurring in 61 quantitative evaluation studies of university-based entrepreneurship education programs. These activities resulted in our Impact Evaluation Research Standards for entrepreneurship educators and researchers.

A key contribution of this study is the development of a descriptive validity framework that we used to review impact studies in entrepreneurship education. This process showed us that while most studies meet baseline reporting standards (e.g., hypotheses, descriptions of research design and analytic methods), there remains a significant opportunity to improve reporting and provide more evidence of internal and external validity. Many studies did not adequately describe the characteristics of interventions, experimental and comparison groups, sample attrition rates, and outcome measures, making it difficult for scholars to judge rigor, generalize findings, or include these studies in meta-analyses or research syntheses, where meaningful conclusions about program effectiveness can be drawn. It is important to remember that descriptive validity is not the key to securing internal and external validity. Instead, its value is in delineating methodological weaknesses and limitations that are inevitable and present in every study, thereby providing a foundation for enhancing the rigor of evaluation research.

In our review, we found that the most common issues preventing entrepreneurship education from meeting contemporary educational evaluation practices and standards were related to research designs, sampling, outcome measures, and analytical techniques. Few studies provided clear descriptions of the study design or employed comparisons between groups or over time. Most relied on convenience sampling instead of randomized or stratified sampling methods. There was minimal evolution in outcome measures, with many remaining imprecise in how they relate to longer-term business creation outcomes. Where there were multiple outcome variables in a single study, statistical corrections for multiple hypothesis testing were rarely made to control for false positives (List et al., 2019). And, while more advanced statistical methods such as HLM and SEM were used, they could not overcome fundamental weaknesses in research design. Interestingly, the prevalence of these issues did not diminish over time, despite ongoing calls for more methodological rigor in entrepreneurship education impact research.

This does not mean, however, that entrepreneurship education is not having a positive impact on students or society. It is clear that widely used measures related to awareness, knowledge, competency, interest, and intent do reflect positive impacts of entrepreneurship education. From an administrative perspective, these measures are very useful when advocating for courses and programs. From a research perspective, scholars argue that these outcomes can influence long-term involvement and success in entrepreneurship, although these links are not well established. It should also be noted that the methodological and reporting issues highlighted in this research are not unique to the field of entrepreneurship, as studies in many academic disciplines lack rigorous educational evaluation methods. It is evident that the scholars whose studies we reviewed spent considerable time developing research questions and hypotheses, designing survey instruments, collecting data, analyzing results, and writing manuscripts. Therefore, omitting elements that could enhance methodological rigor or descriptive validity was probably not intentional. In light of this, our framework and standards can serve as a checklist for educators, researchers, and

journal editors to consider, which can only lead to more rigorous research and reporting that will benefit individual scholars, as well as the field of entrepreneurship education, more generally.

Despite the many hurdles associated with experimental design, it is important to recognize that some studies do approach a "gold standard" for research design and reporting (Henry, 2015), For example, Gielnik et al. (2015) sought "to overcome some shortcomings in previous research" (p. 71) by using a repeated measures experimental design to investigate the long-term effects of action-based training on business creation in an extra-curricular entrepreneurship program in Uganda. They employed a longitudinal design, examining participants 1 month and 12 months post-training. This involved assigning an equal number of volunteers to a treatment and control group referred to as a "waiting group" that received the training after the research period. This allowed the researchers to secure a control group of equal size and background (e.g., interests, motivation, educational levels). This *delayed treatment control group experimental design* offers the potential to obtain more profound evidence of impact by comparing the effect of the intervention over time, between control and treatment groups (i.e., interaction). Nonetheless, the authors acknowledged that even this rigorous study design did not overcome the possibility of self-selection bias.

Several fairly recent longitudinal studies highlight the positive impact of entrepreneurship education using control groups. In a subsequent study, Gielnik et al. (2017) examined short- and long-term training outcomes over a period of 32 months at a university in Kenya. Alaref et al. (2020) examined the impact of entrepreneurship education on students from Tunisian universities 1 year and 4 years after graduation. And, outside of university settings, Elert et al. (2015) studied Swedish participants in a high school program approximately 16 years post-completion, the peak age at which people in Sweden were considered likely to start a business. Although these studies provide valuable information, Singer et al. (2003) indicated that longitudinal designs must have at least three waves of data collection, at meaningful points in time, and outcome measures that change systematically over time. The merit of this approach is that it allows researchers to more profoundly examine the formative processes associated with how interventions influence outcomes over a given period (increase, decrease, constant, or fluctuation). Only Gielnik et al. (2017) adhered to these guidelines and used growth curve modeling to investigate *how* and *when* self-efficacy, passion, and business creation changed, which can offer more precise information to stakeholders wishing to understand, and improve, the impact of their programs.

Since we do not want perfection to get in the way of progress, it is worth repeating that there are also many practical barriers associated with conducting evaluation research in entrepreneurship education that are likely to persist. First, entrepreneurship is fundamentally a career outcome, and the relation of education to a future professional outcome becomes more difficult to measure given the influence of confounding variables over time. Second, there are significant costs and risks associated with conducting experimental and longitudinal research, including the time and resources necessary to track participants, potential low response rates, and consideration of long publishing timelines making it unfeasible for many. Third, pressure to ground research in established constructs makes designing studies able to measure both educational and career outcomes difficult. Fourth, many entrepreneurship faculty members have little training in educational research, and there may be little interest in these types of studies, which historically have not been highly valued in the field, and as such, difficult to publish (Béchard & Grégoire, 2005; Pringle & Michel, 2007).

Nonetheless, it is clear that demonstrating more explicit connections to longer-term career outcomes over time is essential to show stakeholders that entrepreneurship education makes a difference in terms of startup activity and economic development. These outcomes are at the core of the discipline and often serve as the rationale for investing in entrepreneurship programs. They

are also inherent in the constructs we use to measure impact. For example, entrepreneurial intention, which is commonly used as an impact measure for entrepreneurship education among students, has limited value as a concept or construct if it cannot be connected to future entrepreneurial activity through longitudinal research.

More rigorous methods are also essential for publishing in top scholarly journals, from which information about the impact of entrepreneurship education can be disseminated to professional associations, government organizations, donors, funding agencies, and other stakeholders. This requires reporting study details with transparency, not only describing the strengths of the research but also potential threats to validity in terms of overall research design, sample, measures, data, and data analysis. We are confident the descriptive validity framework and Impact Evaluation Research Standards developed in this paper can contribute to improving study design and reporting, by serving as useful guidelines and checklists at many stages of the research process, from the conception of studies to their publication.

Given the priorities and constraints of administrators, educators, and scholars, it is likely that many will continue to focus on shorter-term, education-related measures rather than economic or societal-level impacts. These are practical, accessible, and reflect short-term priorities, such as: What are the characteristics of participating students? Have students met the stated learning objectives? Are they satisfied? What are the contributions of the program to student growth? What is my contribution to learning? How can student learning be improved? And, what is the program's value to the institution? Nevertheless, the findings of this study suggest that significant improvements can be made in measuring these outcomes even more rigorously.

To some degree, without data linking educational outcomes to business creation, it is unclear whether we are even assessing entrepreneurship education in its true sense at all. It is evident that graduates who are able to identify market needs, develop new products, evaluate their financial potential, and advocate for their ideas are increasingly valued by companies big and small. That might be sufficient to advocate for programs at institutions where there is already buy-in and support for entrepreneurship education. However, it limits our ability to examine more closely and empirically how our approaches to teaching entrepreneurship influence career outcomes and how they ultimately impact the economy and society.

Conclusion

Universities are investing in entrepreneurship programs to enable graduates to generate value in the new economy through the founding of new ventures or in the reinvigoration of established organizations. Obtaining objective evidence of impact through more rigorous educational evaluation practices and standards can help educators and scholars verify that entrepreneurship education interventions achieve their desired goals for their targeted populations. This manuscript makes important contributions to the literature by summarizing prior research and examining prevalent research practices to generate a set of methodological standards and best practices for educators and scholars wishing to improve assessment practice through more evidence-based program evaluation and research.

Appendix A

A Summary of Research Synthesis Studies in Entrepreneurship Education.

Study	Method	Goals	*N* of sample studies	Search strategy
Martin et al. (2013)	Meta-analysis	examine the effect of entrepreneurship education and training programs.	42	Restricted to rigorous experimental design
Rideout and Gray (2013)	Systematic review	examine whether entrepreneurship education really works	12	Very restrictive to rigorous impact evaluation studies
Bae et al. (2014)	Meta-analysis	examine the effect of entrepreneurship education on entrepreneurial intention	73	Comprehensive
Nabi et al. (2017)	Systematic review	examine relationships between pedagogical approaches and outcomes	159	Comprehensive
Grégoire et al. (2019)	Systematic review	present an overview of critical validity challenges plaguing entrepreneurship research experiments and assess the validation practices mobilized	144	Somewhat restrictive: Inclusion of at least one experimental manipulation/ treatment designed

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID ID

Nathalie Duval-Couetil https://orcid.org/0000-0002-0260-0208

Notes

1. While attention is being paid to entrepreneurship education and training in many contexts (e.g., secondary and vocational schools, community programs), this study focuses on higher education institutions to minimize variability in program characteristics and target populations.
2. Empirical research finding a higher number of studies using objective measures had different inclusion criteria (see Martin et al., 2013; Nabi et al., 2017).

References

References marked with an asterisk indicate studies included in the meta-analysis.

Aiken, L. S., Stein, J. A., & Bentler, P. M. (1994). Structural equation analyses of clinical subpopulation differences and comparative treatment outcomes: Characterizing the daily lives of drug addicts. *Journal of Consulting and Clinical Psychology*, *62*(3), 488–499. https://doi.org/10.1037/0022-006X.62.3.488

*Akin, H. B., & Demirel, Y. (2015). Entrepreneurship education and perception change: The preliminary outcomes of compulsory entrepreneurship course experience in Turkey. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 34, 15–26. https://hdl.handle.net/20.500.12451/1360

Alaref, J., Brodmann, S., & Premand, P. (2020). The medium-term impact of entrepreneurship education on labor market outcomes: Experimental evidence from University graduates in Tunisia. *Labour Economics*, *62*(4), 101787. https://doi.org/10.1016/j.labeco.2019.101787

*Alcaraz-Rodriguez, R., M. Alvarez, M., & Villasana, M. (2014). Developing entrepreneurial competences in students in the life sciences: The Lifetech Ad-Venture program. *On the Horizon*, *22*(3), 182–191. https://doi.org/10.1108/OTH-11-2013-0053

*Alharbi, J., Almahdi, H., & Mosbah, A. (2018). The impact of entrepreneurship education programmes (EEPs) on the entrepreneurial attitudes among higher education students in Saudi Arabia. *International Journal of Management, Economics and Social Sciences*, *7*(3), 245–271. https://doi.org/10.32327/IJMESS.7.3.2018.16

American Educational Research Association.,American Psychological Association.,National Council on Measurement in Education., &Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*.

*Arranz, N., Ubierna, F., Arroyabe, M. F., Perez, C., & Fdez. de Arroyabe, J. C. (2017). The effect of curricular and extracurricular activities on university students' entrepreneurial intention and competences. *Studies in Higher Education*, *42*(11), 1979–2008. https://doi.org/10.1080/03075079.2015.1130030

*Asghar, M. Z., Gul, F., Seitamaa Hakkarainen, P., & Taşdemir, M. Z. (2019). Validating Entrepreneurial intentions questionnaire to assess the impact of entrepreneurship education. *Ted EĞİtİm Ve Bİlİm*, *44*(197). https://doi.org/10.15390/EB.2019.6105

Azoulay, P., Jones, B., Kim, J. D., & Miranda, J. (2018). Age and High-Growth Entrepreneurship. https://siepr.stanford.edu/system/files/Age%20and%20High%20Growth%20Entrepreneurship%20_%20Integrated.pdf

Bae, T. J., Qian, S., Miao, C., & Fiet, J. O. (2014). The relationship between entrepreneurship education and entrepreneurial intentions: A meta-analytic review. *Entrepreneurship Theory and Practice*, *38*(2), 217–254. https://doi.org/10.1111/etap.12095

*Baidi, B., & Suyatno, B. (2018). Effect of entrepreneurship education, self-efficacy and need for achievement toward students' entrepreneurship intention: Case study in Febi, Iain Surakarta, Indonesia. *Journal of Entrepreneurship Education*, *21*(2), 1–16.

Bauer, K. W. (2004). Conducting longitudinal studies. *New Directions for Institutional Research*, *2004*(121), 75–90. https://doi.org/10.1002/ir.102

Béchard, J. -P., & Grégoire, D. (2005). Entrepreneurship education research revisited: The case of higher education. *Academy of Management Learning & Education*, *4*(1), 22–43. https://doi.org/10.5465/amle.2005.16132536

*Bhat, I. H., & Singh, S. (2018). Analyzing the moderating effect of entrepreneurship education on the antecedents of entrepreneurial intention. *Journal of Entrepreneurship Education*, *21*(1).

Blikstein, P., Kabayadondo, Z., Martin, A., & Fields, D. (2017). An assessment instrument of technological literacies in makerspaces and FabLabs. *Journal of Engineering Education*, *106*(1), 149–175. https://doi.org/10.1002/jee.20156

Boruch, R. F. (1997). *Randomized experiments for planning and evaluation: A practical guide*. Sage.

*Boukamcha, F. (2015). Impact of training on entrepreneurial intention: An interactive cognitive perspective. *European Business Review*, *27*(6), 593–616. https://doi.org/10.1108/EBR-12-2014-0090

Brown, T. A. (2015). *Methodology in the social sciences.Confirmatory factor analysis for applied research* (2nd ed). The Guilford Press.

*Chang, J., & Rieple, A. (2013). Assessing students' entrepreneurial skills development in live projects. *Journal of Small Business and Enterprise Development*, *20*(1), 225–241. https://doi.org/10.1108/14626001311298501

Chen, W., & Tan, J. (2009). Understanding transnational entrepreneurship through a network lens: Theoretical and methodological considerations. *Entrepreneurship Theory and Practice*, *33*(5), 1079–1091. https://doi.org/10.1111/j.1540-6520.2009.00335.x

Cook, T. D., & Campbell, D. T. (1979). The design and conduct of true experiments and quasi-experiments in field settings. In R. T. Mowday & R. M. Steers (Eds.), *Reproduced in part in research in organizations: Issues and controversies*. Goodyear Publishing Company.

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage Foundation.

*Dehghanpour Farashah, A. (2013). The process of impact of entrepreneurship education and training on entrepreneurship perception and intention. *Education + Training*, *55*(8/9), 868–885. https://doi.org/10.1108/ET-04-2013-0053

*DeTienne, D. R., & Chandler, G. N. (2004). Opportunity identification and its role in the entrepreneurial classroom: A pedagogical approach and empirical test. *Academy of Management Learning & Education*, *3*(3), 242–257. https://doi.org/10.5465/amle.2004.14242103

Douglas, E. J., & Shepherd, D. A. (2002). Self-Employment as a career choice: Attitudes, Entrepreneurial intentions, and utility Maximization. *Entrepreneurship Theory and Practice*, *26*(3), 81–90. https://doi.org/10.1177/104225870202600305

*Dutta, D. K., Li, J., & Merenda, M. (2011). Fostering entrepreneurship: Impact of specialization and diversity in education. *International Entrepreneurship and Management Journal*, *7*(2), 163–179. https://doi.org/10.1007/s11365-010-0151-2

Duval-Couetil, N., Gotch, C. M., & Yi, S. (2014). The characteristics and motivations of contemporary entrepreneurship students. *Journal of Education for Business*, *89*(8), 441–449. https://doi.org/10.1080/08832323.2014.933156

Duval-Couetil, N., Shartrand, A., & Reed, T. (2016). The role of entrepreneurship program models and experiential activities on engineering student outcomes. *Advances in Engineering Education*, *5*(1), n1.

Elert, N., Andersson, F. W., & Wennberg, K. (2015). The impact of entrepreneurship education in high school on long-term entrepreneurial performance. *Journal of Economic Behavior & Organization*, *111*(4), 209–223. https://doi.org/10.1016/j.jebo.2014.12.020

*Elmuti, D., Khoury, G., & Omran, O. (2012). Does entrepreneurship education have a role in developing entrepreneurial skills and ventures' effectiveness? *Journal of Entrepreneurship Education*, *15*, 83.

Farrington, D. P. (2003). Methodological quality standards for evaluation research. *The ANNALS of the American Academy of Political and Social Science*, *587*(1), 49–68. https://doi.org/10.1177/0002716202250789

*Fayolle, A., & Gailly, B. (2015). The impact of entrepreneurship education on entrepreneurial attitudes and intention: Hysteresis and persistence. *Journal of Small Business Management*, *53*(1), 75–93. https://doi.org/10.1111/jsbm.12065

*Fayolle, A., Gailly, B., & Lassas-Clerc, N. (2006a). Effect and counter-effect of entrepreneurship education and social context on student's intentions. *Estudios De Economía Aplicada*, *24*(2), 509–524.

*Fayolle, A., Gailly, B., & Lassas-Clerc, N. (2006b). Assessing the impact of entrepreneurship education programmes: A new methodology. *Journal of European Industrial Training*, *30*(9), 701–720. https://doi.org/10.1108/03090590610715022

Fiet, J. O. (2001). The theoretical side of teaching entrepreneurship. *Journal of Business Venturing*, *16*(1), 1–24. https://doi.org/10.1016/S0883-9026(99)00041-5

Frechtling, J., Mark, M. M., Rog, D. J., Thomas, V., Frierson, H., Hood, S., & Hughes, G. *The 2010 User-Friendly Handbook for Project Evaluation.* National Science Foundation. https://www.westat.com/sites/westat.com/files/2010UFHB.pdf

*Fretschner, M., & Weber, S. (2013). Measuring and understanding the effects of entrepreneurial awareness education. *Journal of Small Business Management*, *51*(3), 410–428. https://doi.org/10.1111/jsbm.12019

*Gielnik, M. M., Frese, M., Kahara-Kawuki, A., Wasswa Katono, I., Kyejjusa, S., Ngoma, M., Munene, J., Namatovu-Dawa, R., Nansubuga, F., Orobia, L., Oyugi, J., Sejjaaka, S., Sserwanga, A., Walter, T., Bischoff, K. M., & Dlugosch, T. J. (2015). Action and action-regulation in entrepreneurship: Evaluating a student training for promoting entrepreneurship. *Academy of Management Learning & Education*, *14*(1), 69–94. https://doi.org/10.5465/amle.2012.0107

*Gielnik, M. M., Uy, M. A., Funken, R., & Bischoff, K. M. (2017). Boosting and sustaining passion: A long-term perspective on the effects of entrepreneurship training. *Journal of Business Venturing*, *32*(3), 334–353. https://doi.org/10.1016/j.jbusvent.2017.02.003

Gill, C. E. (2011). Missing links: How descriptive validity impacts the policy relevance of randomized controlled trials in criminology. *Journal of Experimental Criminology*, *7*(3), 201–224. https://doi.org/10.1007/s11292-011-9122-z

Grégoire, D. A., Binder, J. K., & Rauch, A. (2019). Navigating the validity tradeoffs of entrepreneurship research experiments: A systematic review and best-practice suggestions. *Journal of Business Venturing*, *34*(2), 284–310. https://doi.org/10.1016/j.jbusvent.2018.10.002

*Hansemark, O. C. (1998). The effects of an entrepreneurship programme on need for achievement and locus of control of reinforcement. *International Journal of Entrepreneurial Behavior & Research*, *4*(1), 28–50. https://doi.org/10.1108/13552559810203957

*Harms, R. (2015). Self-Regulated learning, team learning and project performance in entrepreneurship education: Learning in a lean startup environment. *Technological Forecasting and Social Change*, *100*, 21–28. https://doi.org/10.1016/j.techfore.2015.02.007

*Hasan, S. M., Khan, E. A., & Nabi, M. N. U. (2017). Entrepreneurial education at university level and entrepreneurship development. *Education + Training*, *59*(7/8), 888–906. https://doi.org/10.1108/ET-01-2016-0020

*Hattab, H. W. (2014). Impact of entrepreneurship education on entrepreneurial intentions of university students in Egypt. *The Journal of Entrepreneurship*, *23*(1), 1–18. https://doi.org/10.1177/0971355713513346

Henry, C. (2015). Entrepreneurship education evaluation: revisiting Storey to hunt for the *heffalump*. *Education + Training*, *57*(8/9), 816–833. https://doi.org/10.1108/ET-05-2015-0035

Jack, S. L., & Anderson, R. (1998). *Entrepreneurship education within the condition of entreprenology*. *Paper presented at the Proceedings of the Conference on Enterprise and Learning, Aberdeen*, Scotland.

*Jahani, S., Babazadeh, M., Haghighi, S., & Cheraghian, B. (2018). The effect of entrepreneurship education on self-efficacy beliefs and entrepreneurial intention of nurses. *Journal of Clinical and Diagnostic Research*, *12*(6). https://doi.org/10.7860/JCDR/2018/31525.11654

*Karimi, S., Biemans, H. J. A., Lans, T., Chizari, M., & Mulder, M. (2016). The impact of entrepreneurship education: A study of iranian students' entrepreneurial intentions and opportunity identification. *Journal of Small Business Management*, *54*(1), 187–209. https://doi.org/10.1111/jsbm.12137

*Karlsson, T., & Moberg, K. (2013). Improving perceived entrepreneurial abilities through education: Exploratory testing of an entrepreneurial self efficacy scale in a pre-post setting. *The International Journal of Management Education*, *11*(1), 1–11. https://doi.org/10.1016/j.ijme.2012.10.001

*Kirby, D. A., & Humayun, H. (2013). Outcomes of an entrepreneurship education programme: An empirical study in Egypt. *International Journal of Management*, *30*(3), 23.

*Kiyani, S. A. (2017). Role of entrepreneurship education on student attitudes. *Abasyn Journal of Social Sciences*, *10*(2), 270–293.

Kolvereid, L. (1996). Organizational employment versus self-employment: Reasons for career choice intentions. *Entrepreneurship Theory and Practice*, *20*(3), 23–31. https://doi.org/10.1177/104225879602000302

Kolvereid, L., & Moen, Ø. (1997). Entrepreneurship among business graduates: Does a major in entrepreneurship make a difference? *Journal of European Industrial Training*, *21*(4), 154–160. https://doi.org/10.1108/03090599710171404

*Lanero, A., Vázquez, J. L., Gutiérrez, P., & García, M. P. (2011). The impact of entrepreneurship education in European universities: An intention-based approach analyzed in the Spanish area. *International Review on Public and Nonprofit Marketing*, *8*(2), 111–130. https://doi.org/10.1007/s12208-011-0067-8

*Lee, S. M., Chang, D., & Lim, S. B. (2005). Impact of entrepreneurship education: A comparative study of the U.S. and Korea. *The International Entrepreneurship and Management Journal*, *1*(1), 27–43. https://doi.org/10.1007/s11365-005-6674-2

List, J. A., Shaikh, A. M., & Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, *22*(4), 773–793. https://doi.org/10.1007/s10683-018-09597-5

Liñán, F., & Chen, Y. –W. (2009). Development and Cross–Cultural application of a specific instrument to measure entrepreneurial intentions. *Entrepreneurship Theory and Practice*, *33*(3), 593–617. https://doi.org/10.1111/j.1540-6520.2009.00318.x

Lösel, F., & Köferl, P. (1989). Evaluation Research on Correctional Treatment in West Germany: A Meta-analysis. In H. Wegener, F. Lösel, & J. Haisch (Eds.), *Criminal behavior and the justice system: Psychological perspectives* (pp. 334–355). Springer Berlin Heidelberg.

*Lourenço, F., & Jayawarna, D. (2011). Enterprise education: The effect of creativity on training outcomes. *International Journal of Entrepreneurial Behavior & Research*, *17*(3), 224–244. https://doi.org/10.1108/13552551111130691

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*(4), 173–181. https://doi.org/10.1023/A:1026595011371

*Mamun, A. A., Nawi, N. B. C., Mohiuddin, M., Shamsudin, S. F. F. B., & Fazal, S. A. (2017). Entrepreneurial intention and startup preparation: A study among business students in Malaysia. *Journal of Education for Business*, *92*(6), 296–314. https://doi.org/10.1080/08832323.2017.1365682

*Maresch, D., Harms, R., Kailer, N., & Wimmer-Wurm, B. (2016). The impact of entrepreneurship education on the entrepreneurial intention of students in science and engineering versus business studies university programs. *Technological Forecasting and Social Change*, *104*, 172–179. https://doi.org/10.1016/j.techfore.2015.11.006

Martin, B. C., McNally, J. J., & Kay, M. J. (2013). Examining the formation of human capital in entrepreneurship: A meta-analysis of entrepreneurship education outcomes. *Journal of Business Venturing*, *28*(2), 211–224. https://doi.org/10.1016/j.jbusvent.2012.03.002

*Matlay, H., Solesvik, M., & Westhead, P. (2014). Cultural factors and entrepreneurial intention. *Education+Training*. https://doi.org/10.1108/ET-07-2014-0075

Meinert, C. L. (1986). *Clinical trials: Design, conduct, and analysis*. Oxford University Press.

Mitchell, M., & Jolley, J. (2001). *Research design explained*. Harcourt College Publishers.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, D., Egger, M., Altman, D. G., & CONSORT. (2012). Consort 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *International Journal of Surgery*, *10*(1), 28–55. https://doi.org/10.1016/j.ijsu.2011.10.001

Morris, M., Kuratko, D., & Cornwall, J. (2013). *Entrepreneurship programs and the modern university*. Edward Elgar Pub.

Nabi, G., Liñán, F., Fayolle, A., Krueger, N., & Walmsley, A. (2017). The impact of entrepreneurship education in higher education: A systematic review and research agenda. *Academy of Management Learning & Education*, *16*(2), 277–299. https://doi.org/10.5465/amle.2015.0026

Neck, H. M., & Greene, P. G. (2011). Entrepreneurship education: Known worlds and new frontiers. *Journal of Small Business Management*, *49*(1), 55–70. https://doi.org/10.1111/j.1540-627X.2010.00314.x

*Nizam Zainuddin, M., & Rozaini Mohd Rejab, M., & Zainuddin, M. N. (2010). Assessing "ME generation's" entrepreneurship degree programmes in Malaysia. *Education + Training*, *52*(6/7), 508–527. https://doi.org/10.1108/00400911011068469

*Ohland, M. W., Frillman, S. A., Zhang, G., Brawner, C. E., & Miller, T. K. (2004). The effect of an entrepreneurship program on GpA and retention. *Journal of Engineering Education*, *93*(4), 293–301. https://doi.org/10.1002/j.2168-9830.2004.tb00818.x

*Oosterbeek, H., van Praag, M., & Ijsselstein, A. (2010). The impact of entrepreneurship education on entrepreneurship skills and motivation. *European Economic Review*, *54*(3), 442–454. https://doi.org/10.1016/j.euroecorev.2009.08.002

*Otache, I. (2019). Enhancing the effectiveness of entrepreneurship education: The role of entrepreneurial lecturers. *Education + Training*, *61*(7/8), 918–939. https://doi.org/10.1108/ET-06-2018-0127

*Passoni, D., & Glavam, R. B. (2018). Entrepreneurial intention and the effects of entrepreneurial education. *International Journal of Innovation Science*, *10*(1), 92–107. https://doi.org/10.1108/IJIS-05-2017-0042

Piperopoulos, P., & Dimov, D. (2015). Burst bubbles or build steam? Entrepreneurship education, entrepreneurial self-efficacy, and entrepreneurial intentions. *Journal of Small Business Management*, *53*(4), 970–985. https://doi.org/10.1111/jsbm.12116

Pittaway, L., & Cope, J. (2007). Entrepreneurship education: A systematic review of the evidence. *International Small Business Journal*, *25*(5), 479–510. https://doi.org/10.1177/0266242607080656

Pittaway, L., Hannon, P., Gibb, A., & Thompson, J. (2009). Assessment practice in enterprise education. *International Journal of Entrepreneurial Behavior & Research*, *15*(1), 71–93. https://doi.org/10.1108/13552550910934468

Plint, A. C., Moher, D., Morrison, A., Schulz, K., Altman, D. G., Hill, C., & Gaboury, I. (2006). Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Medical Journal of Australia*, *185*(5), 263–267. https://doi.org/10.5694/j.1326-5377.2006.tb00557.x

Podsakoff, P. M., MacKenzie, S. B., Lee, J. -Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

*Premand, P., Brodmann, S., Almeida, R., Grun, R., & Barouni, M. (2016). Entrepreneurship education and entry into self-employment among university graduates. *World Development*, *77*(3), 311–327. https://doi.org/10.1016/j.worlddev.2015.08.028

Pringle, C., & Michel, M. (2007). Assessment practices in AACSB-Accredited business schools. *Journal of Education for Business*, *82*(4), 202–211. https://doi.org/10.3200/JOEB.82.4.202-211

*Pruett, M. (2012). Entrepreneurship education: Workshops and entrepreneurial intentions. *Journal of Education for Business*, *87*(2), 94–101. https://doi.org/10.1080/08832323.2011.573594

*Radu Lefebvre, M., & Redien-Collot, R. (2013). "How to do things with words": The discursive dimension of experiential learning in entrepreneurial mentoring dyads. *Journal of Small Business Management*, *51*(3), 370–393. https://doi.org/10.1111/jsbm.12022

*Rauch, A., & Hulsink, W. (2015). Putting entrepreneurship education where the intention to act lies: An investigation into the impact of entrepreneurship education on entrepreneurial behavior. *Academy of Management Learning & Education*, *14*(2), 187–204. https://doi.org/10.5465/amle.2012.0293

Rideout, E. C., & Gray, D. O. (2013). Does entrepreneurship education really work? A review and methodological critique of the empirical literature on the effects of University-Based entrepreneurship education. *Journal of Small Business Management*, *51*(3), 329–351. https://doi.org/10.1111/jsbm.12021

Roman, T., & Maxim, A. (2017). National culture and higher education as pre-determining factors of student entrepreneurship. *Studies in Higher Education*, *42*(6), 993–1014. https://doi.org/10.1080/03075079.2015.1074671

*Rossano, S., Meerman, A., Kesting, T., & Baaken, T. (2016). The relevance of problem-based learning for policy development in University-Business cooperation. *European Journal of Education*, *51*(1), 40–55. https://doi.org/10.1111/ejed.12165

Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2019). *Evaluation: A systematic approach*. Sage publications.

Royse, D., Thyer, B. A., & Padgett, D. K. (2016). *Program evaluation: An introduction to an evidence-based approach*. Cengage Learning.

*Saji, B. S., & Nair, A. R. (2018). Effectiveness of innovation and entrepreneurship education in UAE higher education. *Academy of Strategic Management Journal*, *17*(4).

Schulz, K. F., Altman, D. G., Moher, D., & CONSORT Group. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine*, *8*(1), 18. https://doi.org/10.1186/1741-7015-8-18

Schulz, K. F., Altman, D. G., Moher, D., & Fergusson, D. (2010). Consort 2010 changes and testing blindness in RCTs. *The Lancet*, *375*(9721), 1144–1146. https://doi.org/10.1016/S0140-6736(10)60413-8

*Shealy, K. M., & McCaslan, M. (2018). Incorporating an Entrepreneurial certificate into the pharmacy curriculum. *American Journal of Pharmaceutical Education*, *82*(8), 6701. https://doi.org/10.5688/ajpe6701

*Shinnar, R. S., Hsu, D. K., & Powell, B. C. (2014). Self-efficacy, entrepreneurial intentions, and gender: Assessing the impact of entrepreneurship education longitudinally. *The International Journal of Management Education*, *12*(3), 561–570. https://doi.org/10.1016/j.ijme.2014.09.005

Singer, J. D., Willett, J. B., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press.

*Solesvik, M. Z., Westhead, P., Matlay, H., & Parsyak, V. N. (2013). Entrepreneurial assets and mindsets: Benefit from university entrepreneurship education investment. *Education+ Training*, *55*(8-9), 748–762.

*Souitaris, V., Zerbinati, S., & Al-Laham, A. (2007). Do entrepreneurship programmes raise entrepreneurial intention of science and engineering students? The effect of learning, inspiration and resources. *Journal of Business Venturing*, *22*(4), 566–591. https://doi.org/10.1016/j.jbusvent.2006.05.002

*Stamboulis, Y., & Barlas, A. (2014). Entrepreneurship education impact on student attitudes. *The International Journal of Management Education*, *12*(3), 365–373. https://doi.org/10.1016/j.ijme.2014.07.001

*Sánchez, J. C. (2011b). University training for entrepreneurial competencies: Its impact on intention of venture creation. *International Entrepreneurship and Management Journal*, *7*(2), 239–254. https://doi.org/10.1007/s11365-010-0156-x

Sánchez, J. C. (2013). The impact of an entrepreneurship education program on entrepreneurial competencies and intention. *Journal of Small Business Management*, *51*(3), 447–465. https://doi.org/10.1111/jsbm.12025

SÃ¡, C., & Holt, C. (2019). Profiles of entrepreneurship students: Implications for policy and practice. *Education + Training*, *61*(2), 122–135. https://doi.org/10.1108/ET-06-2018-0139

*Storen, L. A. (2014). Entrepreneurship in higher education. *Education + Training*, *56*(8/9), 795–813. https://doi.org/10.1108/ET-06-2014-0070

Storey, D. J. (2000). Six steps to heaven: Evaluating the impact of public policies to support small business in developed economies. In D. Sexton & H. Landstrom (Eds.), *The Blackwell handbook of entrepreneurship* (pp. 176–193). Blackwell.

*Sun, H., Lo, C. T., Liang, B., & Wong, Y. L. B. (2017). The impact of entrepreneurial education on entrepreneurial intention of engineering students in Hong Kong. *Management Decision*, *55*(7), 1371–1393. https://doi.org/10.1108/MD-06-2016-0392

*Tessema Gerba, D. (2012). Impact of entrepreneurship education on entrepreneurial intentions of business and engineering students in Ethiopia. *African Journal of Economic and Management Studies*, *3*(2), 258–277. https://doi.org/10.1108/20400701211265036

Thursby, J., Fuller, A. W., & Thursby, M. (2009a). US faculty patenting: Inside and outside the University. *Research Policy*, *38*(1), 14–25. https://doi.org/10.1016/j.respol.2008.09.004

*Thursby, M. C., Fuller, A. W., & Thursby, J. (2009b). An integrated approach to educating professionals for careers in innovation. *Academy of Management Learning & Education*, *8*(3), 389–405.

Uman, L. S. (2011). Systematic reviews and meta-analyses. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *20*(1), 57–59.

*Varamäki, E., Joensuu, S., Tornikoski, E., & Viljamaa, A. (2015). The development of entrepreneurial potential among higher education students. *Journal of Small Business and Enterprise Development*, *22*(3), 563–589. https://doi.org/10.1108/JSBED-02-2012-0027

*von Graevenitz, G., Harhoff, D., & Weber, R. (2010). The effects of entrepreneurship education. *Journal of Economic Behavior & Organization*, *76*(1), 90–112. https://doi.org/10.1016/j.jebo.2010.02.015

*Vorley, T., & Williams, N. (2016). Not just dialling it in. *Education + Training*, *58*(1), 45–60. https://doi.org/10.1108/ET-01-2015-0010

*Walter, S. G., Parboteeah, K. P., & Walter, A. (2013). University departments and self–employment intentions of business students: A cross–level analysis. *Entrepreneurship Theory and Practice*, *37*(2), 175–200. https://doi.org/10.1111/j.1540-6520.2011.00460.x

*Westhead, P., & Solesvik, M. Z. (2016). Entrepreneurship education and entrepreneurial intention: Do female students benefit? *International Small Business Journal: Researching Entrepreneurship*, *34*(8), 979–1003. https://doi.org/10.1177/0266242615612534

Wiklund, J., Nikolaev, B., Shir, N., Foo, M. -D., & Bradley, S. (2019). Entrepreneurship and well-being: Past, present, and future. *Journal of Business Venturing*, *34*(4), 579–588. https://doi.org/10.1016/j.jbusvent.2019.01.002

Yi, S., & Duval-Couetil, N. (2018). What drives engineering students to be entrepreneurs? Evidence of validity for an entrepreneurial motivation scale. *Journal of Engineering Education*, *107*(2), 291–317. https://doi.org/10.1002/jee.20199

*Zhang, Y., Duysters, G., & Cloodt, M. (2014). The role of entrepreneurship education as a predictor of university students' entrepreneurial intention. *International Entrepreneurship and Management Journal*, *10*(3), 623–641. https://doi.org/10.1007/s11365-012-0246-z

Author Biographies

Soohyun Yi is an Assistant Professor of Educational Psychology at Texas Tech University, Lubbock, TX. Her research focuses on student motivation, learning trajectories, and the impact of educational programs.

Nathalie Duval-Couetil is a Professor in the Department of Technology Leadership and Innovation at Purdue University, where she also directs cross-campus entrepreneurship initiatives. Her research focuses on entrepreneurship education, evaluation and assessment, and gender issues.