

Measuring impact through experimental design in entrepreneurship education: A literature review and research agenda

Authors:

Kjersti Kjos Longva

Norwegian University of Science and Technology, Norway

Lene Foss

UiT – The Arctic University of Norway, Norway

Corresponding author:

Kjersti Kjos Longva, Department of International Business, Norwegian University of Science and Technology, Post Box 1517, NO-6025 Aalesund, Norway. Email: kjersti.kjos.longva@ntnu.no

Acknowledgements

We express our gratitude to Professor Øivind Strand, Professor Åsa Lindholm Dahlstrand and Professor Henry Colette for insightful comments on earlier versions of this paper. We also thank the anonymous referees for their valuable comments.

ABSTRACT:

True experimental design and quasi-experimental design are considered to be rigorous research designs appropriate for assessing the impact of pedagogical interventions. This study explores the extent and application of experimental design in impact research on entrepreneurship education (EE) based on a systematic literature review. The findings reveal a substantial lack of methodologically rigorous studies on EE impact, which has severe implications for the accumulated knowledge on the subject. Furthermore, the article summarizes the findings from the body of experimental impact studies with a strong research design and concludes by indicating fruitful avenues for future research.

KEYWORDS: Entrepreneurship education, education impact, entrepreneurial learning outcomes, experimental design, systematic literature review

INTRODUCTION

Entrepreneurship is recognized as an important driver of economic growth (Audretsch et al., 2006). There has, consequently, been an increasing propensity in government policy to promote entrepreneurship education (EE) as a means of stimulating economic growth (Martinez et al., 2010; O'Connor, 2013). The introduction and development of EE courses demand substantial investments, in terms of both time and money, from faculty, educational institutions, sponsors, policymakers and other stakeholders. It is accordingly important to understand the impact that EE can have on students: for example, whether they develop an entrepreneurial mindset through such courses, or whether EE actually contributes to increased start-up rates after graduation.

There has been substantial growth in impact research on EE as stakeholders seek to understand its consequences for students and society as a whole (Martin et al., 2013; Blenker et al., 2014; Bae et al., 2014; Nabi et al., 2016a). However, empirical research has produced rather mixed results on the impact of EE using various measures of entrepreneurial outcome (Martin et al., 2013; Lorz et al., 2013; Bae et al., 2014). While some scholars have found a positive impact on, for instance, entrepreneurial attitudes and intentions (Kolvereid and Moen, 1997; Fayolle et al., 2006; Wilson et al., 2007) and entrepreneurial behaviour (Kolvereid and Moen, 1997; Lange et al., 2011; Elert et al., 2015), others have obtained mixed results (Souitaris et al., 2007; Oosterbeek et al., 2010). Some have even found indications of a negative impact on entrepreneurial orientation (Mentoor and Friedrich, 2007) and entrepreneurial intention (Oosterbeek et al., 2010; Von Graevenitz et al., 2010). Therefore, how EE affects students, and via which mechanisms, remains unexplained.

The growing body of impact studies on EE has, therefore, received considerable criticism. A major concern has been the lack of empirical studies that are methodologically robust (Martin et al., 2013; Bae et al., 2014; Fayolle and Linan, 2014), a weakness that has also been highlighted in research on management education in general (Rynes and Brown, 2011; Köhler et al., 2017). Köhler et al. (2017) argue that, to gain legitimacy for a field and publish impactful research, impact studies need to be designed in a way that provides strong evidence for such effects. Rigorous experimental design is, according to Slavin (2002: 18), 'the design of choice for studies that seek to make causal conclusions, and particularly evaluation of education innovations' and ought to be the preferred choice when addressing educational impact (Johnson and Christensen, 2012). In this study, we define rigorous or strong experimental design as true experiments or quasi-experiments that make use of a longitudinal design (as opposed to a cross-sectional design) and have control groups for comparison (Cook and Campbell, 1979; Shadish et al., 2002). Accordingly, these would be suitable research designs for studying the impact of EE as a pedagogical intervention. The degree to which strong experimental design is actually applied in EE impact research is, however, not known, although EE impact research has been criticized for reporting impact without the necessary level of methodological

rigour. This can have severe implications for the accumulated knowledge about impact in EE research, on which educators and policymakers have to base their actions. Thus it is critical to establish a strong experimental design for EE impact research when providing stakeholders with empirical evidence about the relationship between EE and entrepreneurial learning outcomes.

Based on the above, we believe that the use of experimental research design in EE impact research requires further investigation. The twin objectives of this systematic literature review (SLR) are, therefore, 1) to explore the diffusion of experimental impact studies in EE research and the extent to which those studies have a strong experimental research design (i.e. apply a true experimental design or a quasi-experimental design) and 2) to synthesize the findings on entrepreneurial outcome measures in studies with a strong experimental research design.

To address these objectives, we use an SLR approach to explore published research reported in 65 journals listed by the Association of Business Schools (ABS). By applying established categories of experimental research design, we are able to classify quantitative EE impact studies according to the robustness of their research design and to provide an overview of the status quo in EE impact research. While our review highlights examples in which experimental research design has been applied successfully, it also sheds light on the scarcity of strong experimental design in EE impact studies and the threat this poses for the reliability of previous empirical findings. Furthermore, we provide a synthesis of empirical studies with strong experimental research design in order to establish the cumulative knowledge in EE that can be traced back to methodologically robust quantitative studies. Our study contributes to EE scholarship from both methodological and theoretical perspectives by furthering our understanding of the use of experimental research design in EE impact studies. We propose key avenues for further research that hold the potential to strengthen and build legitimacy for the field of EE research, and the findings from the study should be of value to scholars applying experimental design in their empirical work, as well as practitioners and policymakers who are seeking to better understand the impact of EE as a pedagogical intervention.

The content of the rest of the paper is as follows. The next section addresses the use of EE outcome measures and outlines findings in earlier reviews and meta-analyses of EE. Thereafter, the methodological approach is presented along with a recap of seminal contributions on experimental research design to draw up experiment classifications. Next, the descriptive and qualitative findings of the SLR are reported, and then the paper concludes with a discussion of the findings, our conclusions and the implications of our work for future research on EE.

RESEARCH CONTEXT: MEASURING THE IMPACT OF EE

Impact studies on EE aim to establish whether a pedagogical intervention has caused any change in specific outcome variables. The outcomes measured need to be carefully aligned with the intended learning outcomes for the EE course (Kamovich and Foss, 2017) and may address changes in students' hearts, minds and behaviour (Souitaris et al., 2007). The importance of evaluating the outcomes of EE has been widely acknowledged (Mets et al., 2017), and different frameworks have been suggested for categorizing entrepreneurship learning outcomes. Fisher et al. (2008) developed a tripartite framework drawing on seminal contributions in the education literature (Bloom et al., 1956; Kraiger et al., 1993) which categorizes entrepreneurial learning outcomes as cognitive, skill-based or affective. Cognitive outcomes refer to knowledge, comprehension and critical thinking about entrepreneurship; skill-based outcomes are linked to the skills necessary to start a business; and affective outcomes comprise entrepreneurial attitudes, volition and behavioural preferences.

An alternative framework for teaching and learning entrepreneurship was suggested by Kyrö (2008). The framework consists of three constructs: cognition, affection and conation. Compared with the framework of Fisher et al. (2008), skill-based learning outcomes do not comprise a separate category, but rather are included in cognitive learning outcomes. Furthermore, affective learning outcomes are divided into affection and conation. While affection refers to emotions and perceptions, conation takes the mind one step closer to behaviour, as it describes how one acts on thoughts and feelings via impulse or directed effort (Ajzen, 1989).

Four EE outcomes drawn from the above sources are shown in Table 1, along with behavioural outcomes as a fifth category. After all, developing cognitive, skill-based, affective and conative entrepreneurial outcomes should ultimately lead to entrepreneurial behaviour and socio-economic outcomes in real life; for example, through employability, business creation, intrapreneurship or social entrepreneurship (Kozlinska, 2016; Mets et al., 2017). Hence it is essential to establish an understanding of the impact of EE in all five outcome categories of EE impact research.

Table 1. Categories of outcome measures in EE impact studies.

Outcome measure	Examples of constituents
Cognitive	Knowledge: comprehension about entrepreneurship; business basics Traits: need for achievement, proactiveness, self-esteem, risk propensity
Skill-based	Business modelling; opportunity recognition; creative thinking; teamwork
Affective	Passion/inspiration; attitude to entrepreneurship; subjective norm
Conative	Entrepreneurial intention; entrepreneurial self-efficacy
Behavioural	Nascency; venture creation; intrapreneurship; social entrepreneurship; employability

There have been several previous attempts to summarize findings on EE impact through SLRs and meta-analyses. In 2007, Pittaway and Cope reviewed 184 papers published between 1970 and 2004 in an SLR, and concluded that EE appeared to have an impact on student propensity and intentionality towards entrepreneurship. They emphasized that there was a lack of research on whether EE actually led to entrepreneurial behaviour and, more specifically, on the link between different forms of pedagogy and student entrepreneurial outcomes. Their findings are supported by Mwasalwiba (2010), who in his literature review also highlights the substantial focus on attitudes and intentions, and the failure to link these to actions. He further calls for broader outcome definitions.

A positive impact on skills and knowledge, attitudes, intentions and nascent entrepreneurship is also acknowledged in SLRs by Rideout and Gray (2013) and Lorz et al. (2013). Both reviews draw attention to the methodological weaknesses and deficiencies found in most EE impact studies. This tendency is further confirmed in two meta-analyses on EE by Martin et al. (2013) and Bae et al. (2014). Using human capital as a theoretical lens in a meta-analysis of 42 studies, Martin et al. (2013) found a significant positive association between EE/training and entrepreneurial human capital, as well as between EE/training and entrepreneurship outcomes. Closer examination of the findings did, however, reveal that studies without a strong experimental design tended to overestimate the positive association. When studies with pre- and post-measurement and control groups were isolated, the effect size was substantially reduced.

Bae et al. (2014) report similar findings on how methodological rigour influences empirical findings on EE. Their meta-analysis of 73 studies found a small significant correlation between EE and entrepreneurial intention. However, after controlling for the intentions that students had before EE, the association was no longer significant. Hence, when controlling for self-selection bias by introducing pre-intervention measurement, the actual impact of EE becomes unclear. Bae et al. (2014) further established the role of cultural values as moderators in the relationship between EE and entrepreneurial intention.

A recent SLR by Nabi et al. (2016a) of 159 impact studies of EE in higher education also recognizes that there are substantial methodological weaknesses in those studies. However, their main critique concerns the outcome measures and the lack of detail on pedagogical intervention studies. The authors argue that there is too much focus on short-term subjective impact measures as opposed to long-term behavioural measures such as venture creation and performance. They also lobby for novel impact indicators related to, for example, affective measures such as emotion and mindset. Furthermore, in line with Martin et al. (2013), they call for more research to explain the contradictory findings of impact studies, for instance by including person-, context- and model-specific moderators.

Thus, despite the increasing body of impact studies on EE, it appears that we still have scant knowledge on this matter. While there are several insightful indications about impact and outcomes

in existing empirical studies, there are also rather ambiguous findings that require further investigation. Hence, in the remainder of this paper, we first set out to explore the application of experimental research design in EE impact research. Subsequently, empirical studies with a strong experimental design are examined to establish what can actually be considered reliable knowledge about the impact of EE as a pedagogical intervention.

A SYSTEMATIC LITERATURE REVIEW APPROACH

This study is based on an SLR approach, which aims to make the literature search and review process transparent and replicable. According to Pittaway and Cope (2007) and Nabi et al. (2016a), SLRs have become a well-established methodological approach in the fields of both entrepreneurship and EE and are especially valuable when attempting to sum up evidence over long periods. Figure 1 documents the different stages of our SLR process, for which the starting point was our research objectives: first, to identify experimental impact studies on EE and, subsequently, to review extant knowledge on EE impact produced by rigorous studies with a strong experimental design.

Our SLR is based on a journal-led search in selected peer-reviewed journals. While admittedly this approach may have certain limitations in terms of potentially excluding relevant articles outside the selected journals, it was necessary to ensure the feasibility of the SLR by generating hundreds rather than thousands of hits. It was also essential to target high-quality and impactful EE research; hence we followed Blenker et al. (2014) and Wang and Chugh (2014) in applying the ABS Academic Journal Quality Guide to identify journals, as the Guide provides an indication of the quality and impact of the scientific contribution of articles included in the listed journals. As EE is a research field at the interface between entrepreneurship and business and management education, the literature search included all journals in the ABS subject areas ‘entrepreneurship’ and ‘management development and education’. The journal searches were conducted using the databases Science Direct, Elsevier Scopus, ABI Inform and Business Source Complete for articles published up to and including December 2017. Journals that were not accessible through the databases were searched manually. Titles, abstracts and keywords were searched using the primary Boolean search term (‘entrepreneurship education’ OR ‘enterprise education’), and the secondary search term (‘impact’ OR ‘effect’ OR ‘outcome’ OR ‘learning’) was used for a full-text search to identify quantitative impact studies on EE. The first database search, after the removal of duplicates, resulted in 613 articles.

Subsequently, we reviewed titles, abstracts and the methodology sections of the articles, and excluded those that did not meet the inclusion criteria for quantitative impact studies described in Figure 1. This process left 132 articles. While SLRs have advantages over traditional ad hoc narrative

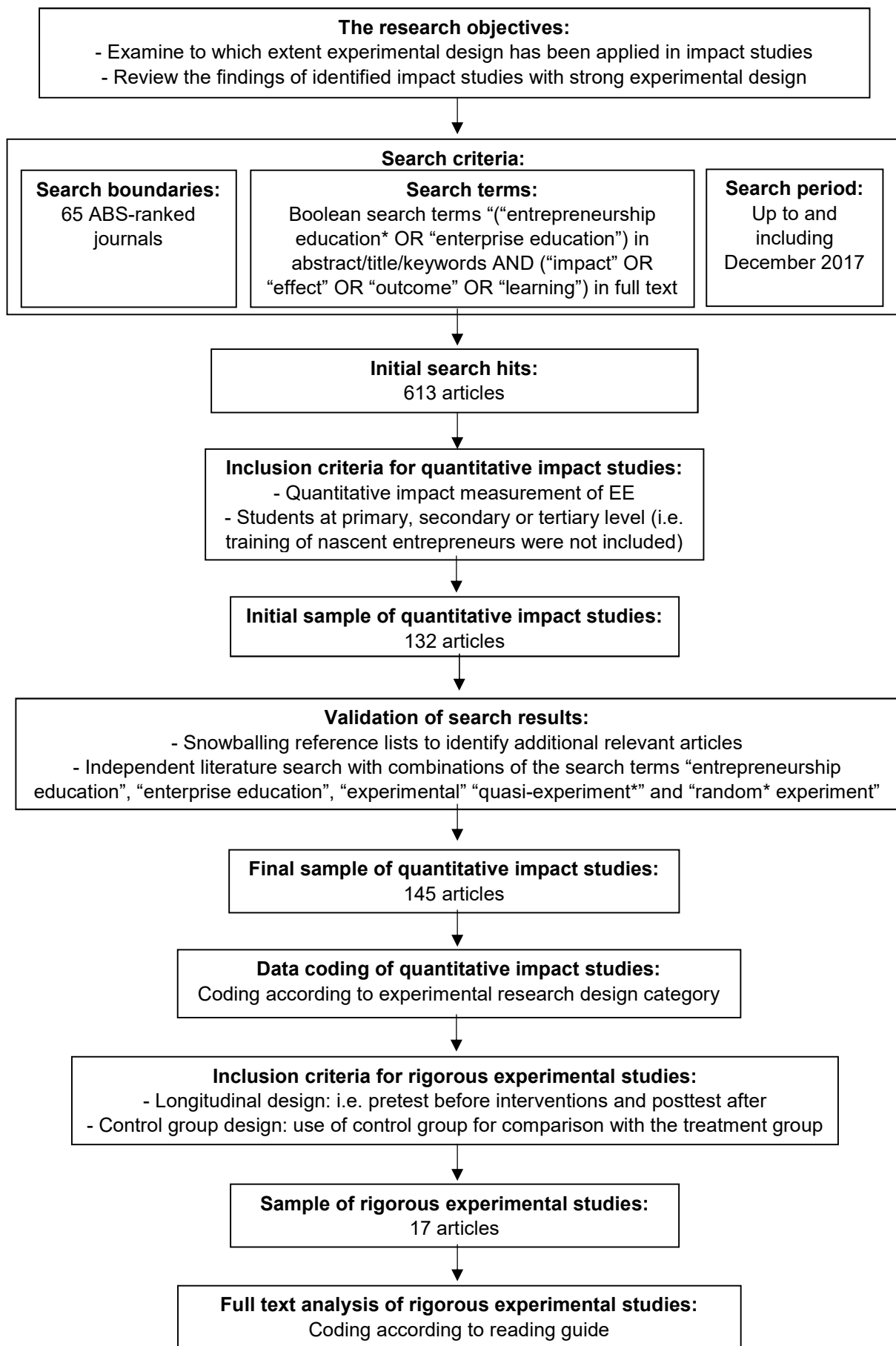


Figure 2. Stages in the SLR process

reviews in that they provide a set of clear steps to systematically generate evidence (Tranfield et al., 2003), a potential drawback is the risk of excluding relevant articles. Hence, as an additional measure to validate the search results and ensure that relevant publications had not been overlooked, we conducted independent literature searches. We also applied snowballing to identify other relevant ABS journals by searching the reference list of the other identified articles. Through this process we expanded our search to include in addition the *European Economic Review* and the *Journal of Economic Behavior & Organization*, which are included in the subject area ‘economics, econometrics and statistics’ in the ABS list.

After validation of the SLR search results, the final sample consisted of 145 articles that met the inclusion criteria for quantitative impact studies. These were coded according to the experimental research design category as described in the following section, and a subgroup of 17 articles that could be classified as rigorous experimental studies with a strong research design were accordingly subjected to a full-text analysis.

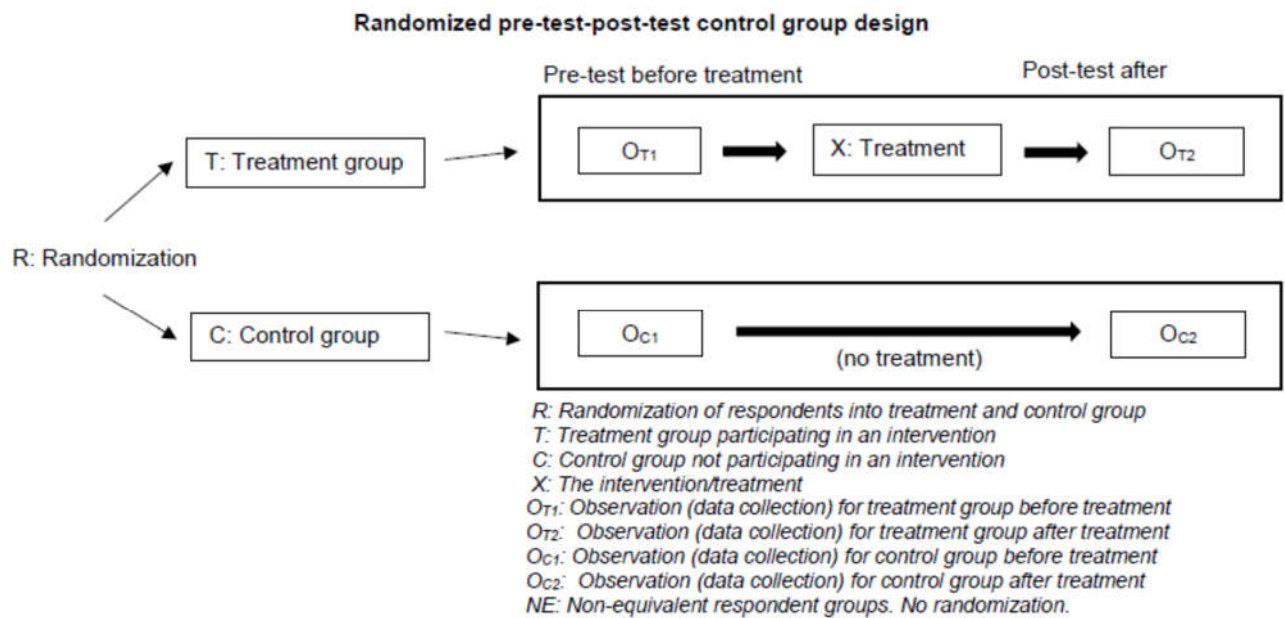
Analysis

Drawing on Blenker et al. (2014) and Wang and Chugh (2014), amongst others, we constructed a thematic reading guide for reviewing and coding the articles (see Appendix). The 145 quantitative studies were coded according to general information (author(s), year, title and journal) and the type of experimental design. If a study was classified as being either a true experiment or a quasi-experiment, it was further coded in accordance with the remainder of the reading guide by focusing on the outcome variables utilized and recording contextual variables stated in the studies, such as the characteristics of pedagogical intervention, sample characteristics and time frame.

The SLR applies three categories of experimental design following Cook and Campbell (1979) and Shadish et al. (2002): true experimental, quasi-experimental and pre-experimental design. Within these three categories, there are various types of experimental design. The ones that were used for coding impact studies in this SLR are depicted in Figure 2.

Experimental designs differ with respect to three characteristics: 1) whether the experiment makes use of control groups; 2) whether randomization into treatment and control groups is applied; and 3) whether the research design is longitudinal as opposed to cross-sectional. The upper half of Figure 2 illustrates the classic true experiment – the randomized pre-test–post-test control group design, in which all three of the above characteristics are present. Here, participants are randomly assigned to either a control group, C, or a treatment group, T, and thereafter are given a pretest O_{T1} or O_{C1} to ensure that the groups do not differ from the outset. Thereafter, T undergoes treatment X (for example, in the form of an EE course), while C does not take part in the course. Afterwards, a post-

test O_{T2} or O_{C2} is completed, and any difference between T and C is assumed to be due to X. The lower half of Figure 2 exemplifies the design of the randomized pre-test–post-test control group design, together with other experimental designs relevant to EE impact research.¹



True randomized experiments (TED)	Quasi-experimental design (QED)	Pre-experimental design (PED)
<i>Randomized pretest-posttest control group design:</i> R O _{T1} X O _{T2} R O _{C1} O _{C2}	<i>Non-equivalent pretest-posttest control group design:</i> NE O _{T1} X O _{T2} NE O _{C1} O _{C2}	<i>One-group posttest only design:</i> - X O _{T2}
<i>Randomized Solomon 4-group design:</i> R X O _{T2} R O _{C2} R O _{T1} X O _{T2} R O _{C1} O _{C2}		<i>Non-equivalent posttest only design:</i> NE X O _{T2} NE O _{C2}
		<i>One-group pretest-posttest design</i> O _{T1} X O _{T2}

Figure 2. Types of experimental research design as described by Cook and Campbell (1979).

¹ For an in-depth discussion of the various experimental designs, see, for example, Campbell and Stanley (1963), Cook and Campbell (1979) and Shadish et al. (2002).

The reason for making use of control groups, randomization of participants and longitudinal design is to control for confounding variables that threaten internal validity. As the key objective of an impact study of education is to find evidence of a causal link between the education intervention and the observed outcomes, it is advisable to apply strong experimental research that controls for confounding variables and, thereby, to exclude alternative explanations and rival hypotheses for observed effects (Mertens, 2010; Johnson and Christenson, 2012). According to Johnson and Christenson (2012), true experimental and quasi-experimental designs could consequently be considered strong experimental designs, while a pre-experimental design is characterized as a weak experimental design. The presence of randomization, control groups and longitudinal design in true experimental designs controls for confounding variables such as *history* (when environmental events during an experiment influence the dependent variable), *maturity* (biological or psychological changes during an experiment due to the passage of time), *testing* (participants becoming test-wise post-test due to earlier pre-tests), *mortality* (participant drop-out during an experiment), *statistical regression* (when diverging scores of extreme groups regress towards the mean when testing is repeated), and *selection* (systematic differences between treatment and control groups due to self-selection) (Campbell and Stanley, 1963; Cook and Campbell, 1979). The randomized pre-test–post-test control group design and the randomized Solomon four-group design² shown in Figure 2 are accordingly considered to be strong experimental designs as they apply randomization, control groups and longitudinal design (Shadish et al., 2002), and findings based on a true experimental design would consequently provide strong evidence of causal links between EE courses and entrepreneurial learning outcomes.

In many educational real-life settings, random assignment is not a realistic option. Following Cook and Campbell (1979), the quasi-experiment would then be the recommended design. The non-equivalent pre-test–post-test control group design is the most relevant quasi-experimental design in EE impact studies, as it enables comparison of EE and non-EE students. In this case, students attending an EE course would constitute the treatment group. The control group would comprise students not attending an EE course, but otherwise would be as similar to the student treatment group as possible. Without randomization, the internal validity of the design faces challenges in terms of selection, maturation, history and statistical regression (Shadish et al., 2002). Nonetheless, with the presence of both control groups and a longitudinal design, it can still be considered a strong experimental design with which it is reasonable to claim causality between an EE course and observed outcomes.

Pre-experimental designs are considered to be weaker experimental research designs due to their limited control of potentially confounding variables (Shadish et al., 2002; Johnson and Christenson, 2012). The one-group post-test only design is considered to be the weakest among these

alternatives. With this research design, students attending an EE course would take a post-test after finishing it. The design poses many threats to internal validity and has been referred to by Campbell and Stanley (1963: 5) as having ‘...such a total absence of control as to be of almost no scientific value’. The design is subject to threats of history, maturation and mortality as it does not apply neither a control group nor a pre-test. The non-equivalent post-test only design introduces comparison groups and the one-group pre-test–post-test design makes use of measurements before and after EE interventions. However, both research designs still face basic problems due to threatened internal validity. Thus, relying on a pre-experimental design when attempting to address the impact of EE courses can be problematic in terms of claiming causality. Therefore, a true experimental design or a quasi-experimental design should be the preferred alternative in quantitative impact studies on EE, and the following section presents the degree to which these rigorous experimental designs are being applied in EE impact studies.

FINDINGS

Descriptive analysis

As noted above, the systematic search in ABS-listed journals resulted in 145 identified quantitative impact studies on EE. Figure 3 shows the journals in which these were published. The figure identifies two major outlets for quantitative impact studies on EE: *Education and Training*, which has published 38 articles, and *Industry and Higher Education*, with 20 published quantitative impact studies on EE.

The coding of the 145 quantitative impact studies revealed that only 17 articles were experimental studies with a strong design; i.e. a true experimental design (TED) or a quasi-experimental design (QED). The remaining 128 quantitative impact studies were described as having a weak pre-experimental design (PED) (see Figure 4). Among the studies, 28% had the weakest of the pre-experimental designs, the one-group post-test only design, while 28% had the non-equivalent post-test only design, and 32% had a one-group pre-test–post-test design. Among the 17 experimental studies, four had a true experimental design, while there were 13 quasi-experimental studies with a non-equivalent pre-test–post-test control group design. Hence, the analysis showed that only 11.7% of the quantitative impact studies met the standards for a strong experimental design.

Figure 5 illustrates the increased amount of quantitative impact studies in recent decades, and depicts the rather limited application of experimental design in comparison. Especially in the last 10

² The Solomon four-group design was developed to overcome threats of testing in pre-test–post-test design, as the two extra control groups allow researchers to test whether the pre-test itself has an impact on the participants (Cook and Campbell, 1979).

years, there has been a considerable yearly increase in the amount of impact studies. There has, however, not been corresponding growth in impact studies with a strong experimental design.

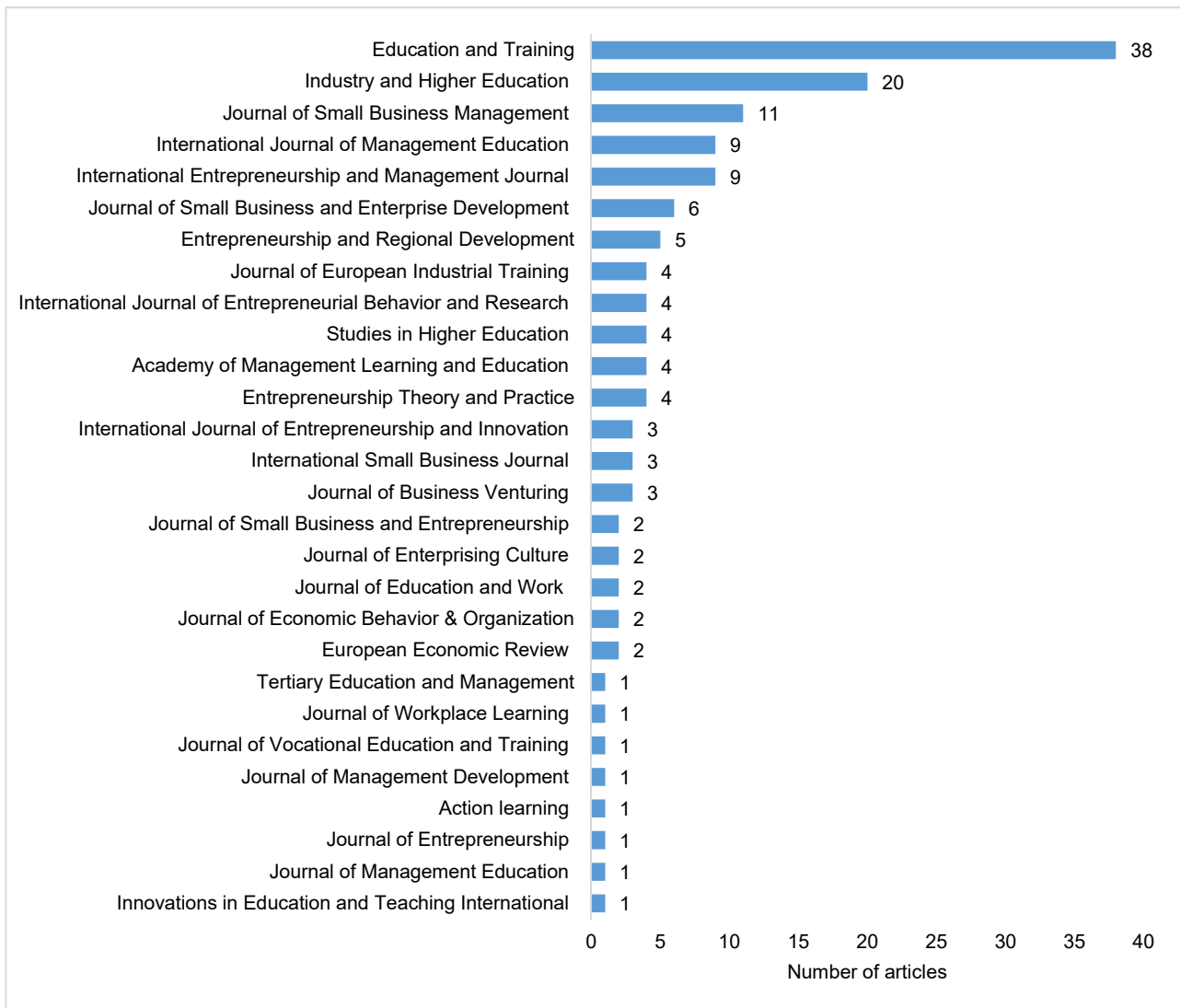


Figure 3. Overview of ABS-listed journals that have published EE impact studies ($n = 145$)

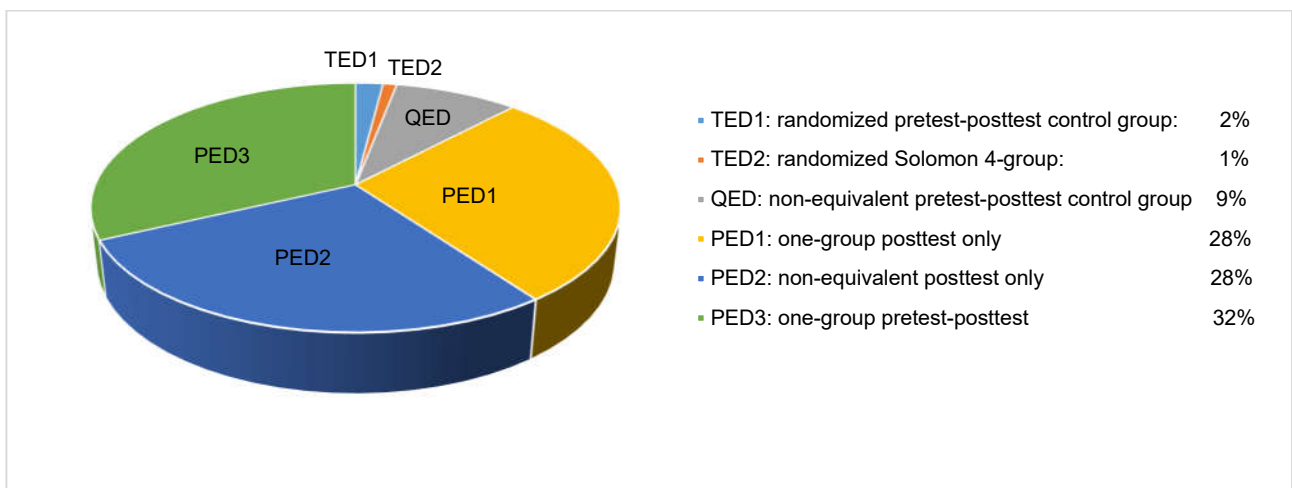


Figure 4. Types of experimental design in EE impact studies ($n = 145$)

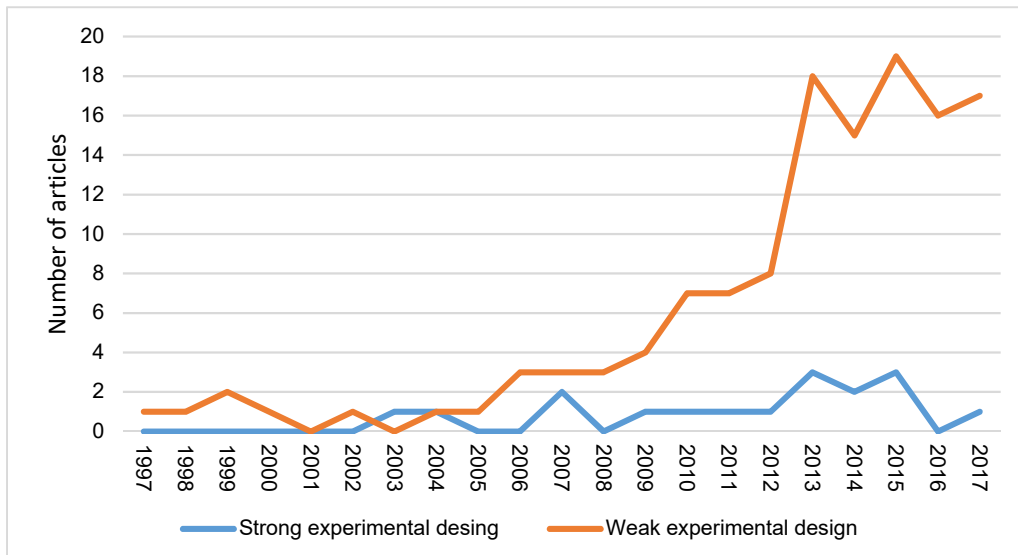


Figure 5. Twenty-one years of quantitative impact studies (1997-2017; $n = 145$)

The descriptive findings therefore point towards considerable challenges for impact research on EE. On a positive note, the amount of EE impact studies is increasing and there are high-quality journals in which this discussion is taking place. Nevertheless, the rigour of the research design is a substantial issue when building accumulated knowledge in the field. When only 11.7% of quantitative impact studies apply a strong experimental design, this has severe implications in terms of making inferences about EE impact.

Qualitative analysis

Entrepreneurial outcome measures. The findings from the analysis of the 17 identified studies applying a strong experimental design illustrate how conative outcomes in terms of entrepreneurial self-efficacy/feasibility and entrepreneurial intention are the most frequently applied outcome measure (Table 2). Of the 17 studies, 13 use either one or both of these as outcome variables. Cognitive outcomes such as knowledge and traits (six studies), as well as skill-based outcomes (seven studies), have also received attention. In terms of affective outcomes, seven studies apply attitude to entrepreneurship as an outcome variable, while subjective norm and passion/inspiration have received less attention. In fact, only two studies (Souitaris et al., 2007; Varamäki et al., 2015) make use of subjective norm to measure EE effect, while only Nabi et al. (2016b) and Gielnik et al. (2017) have recently addressed impact on entrepreneurial inspiration and entrepreneurial passion, respectively. With regard to actual entrepreneurial behaviour, the impact on nascency has been examined in only three studies (Karlsson and Moberg, 2013; Gielnik et al., 2015; Rauch and Hulsink, 2015), while

actual venture creation remains almost unaddressed, with one honourable exception (Gielnik et al., 2017).

SEE TABLE 2 AT THE END OF THIS PAPER

Although the majority of the 17 studies report a positive impact on the various outcome measures, the findings are still mixed – see Table 3 for a summary.³ In terms of entrepreneurial knowledge, Volery et al. (2013), Gielnik et al. (2015) and Nabi et al. (2016b) find a positive impact of EE, while Huber et al. (2014) find no significant relationship. The findings are also mixed with regard to entrepreneurial traits. While Huber et al. (2014) report a positive impact on traits such as need for achievement, social orientation and proactivity, studies by Mentoor and Friedrich (2007), Oosterbeek et al. (2010) and Volery et al. (2013) all report non-significant impacts on traits such as the need for achievement, the need for autonomy, the need for power, endurance, risk propensity and innovation propensity.

Outcome measure		Positive	Non-significant	Negative
Cognitive	Knowledge	3 (n=542)	1 (n=1727)	
	Traits	1 (n=1727)	3 (n=813)	1 (n=418)
Skill-based	Skills	7 (n=3015)	2 (n=395)	
	Passion/inspiration	2 (n=214)		
Affective	Attitude to entrepreneurship	5 (n=713)	2 (n=190)	1 (n=66)
	Subjective norm	1 (n=124)	1 (n=43)	1 (n=23)
Conative	Entrepreneurial intention	5 (n=1099)	3 (n=446)	3 (n=1897)
	Feasibility	9 (n=3278)	2 (n=190)	
Behavioral	Nascency	2 (n=224)	1 (n=51)	
	Venture creation	2 (n=287)		

Table 3. Finding on outcome measures.

³ Some studies use multiple outcome measures and their sample can therefore be found more than once in Table 3.

The impact on skills is, however, mainly positive, and EE is reported to have a positive impact on opportunity identification and exploitation (DeTienne and Chandler, 2004; Thursby et al., 2009; Volery et al., 2013), proactiveness and risk-taking (Sanchez, 2011, 2013; Huber et al., 2014), and analysing, motivating and creativity (Huber et al., 2014). However, Oosterbeek et al. (2010) report non-significant results on entrepreneurial skills.

The studies on entrepreneurial attitude are, with two exceptions (Souitaris et al. (2007; Varamäki et al., 2015), overwhelmingly positive regarding the impact of EE. Studies on other affective outcome measures, however, remain scarce. Nevertheless, two recent studies report a positive impact on entrepreneurial passion (Gielnik et al., 2017) and entrepreneurial inspiration (Nabi et al., 2016b), while Souitaris et al. (2007) establish a positive impact on subjective norm, in contrast to the non-significant and negative findings of Varamäki et al. (2015).

With regard to conative outcomes, nine studies report a positive impact on feasibility/perceived behavioural control/entrepreneurial self-efficacy. Souitaris et al. (2007) and Varamäki et al. (2015) are the only studies that present non-significant findings. The most equivocal results derive from the studies that address entrepreneurial intention: five report a positive impact, two found no significant difference (Volery et al., 2013; Nabi et al., 2016b), one found both non-significant and negative impacts depending on the pedagogics (Varamäki et al., 2015), and two even found a purely negative impact (Oosterbeek et al., 2010; Huber et al., 2014). By far the largest sample size is to be found in the study by Huber et al. (2014). Therefore, when summarizing the samples and results, we find the following distribution of EE impact on entrepreneurial intention: positive impact (n = 1099); non-significant impact (n = 446); and negative impact (n = 1897). Accordingly, although it is the most frequently applied outcome measure in impact studies, evidence of the actual impact of EE on entrepreneurial intention remains highly inconclusive.

Studies on actual entrepreneurial behaviour signal positive findings about entrepreneurial nascency (Gielnik et al., 2015; Rauch and Hulsink, 2015) and new venture creation (Gielnik et al., 2015; 2017). There is, however, a sample size issue here as the studies on nascency had a total sample size of only 224 and the studies on venture creation had a total sample size of 287.

Therefore, although the majority of studies report positive impacts, there are also several with non-significant findings and some even with a negative impact. Consequently, it becomes difficult to conclude anything on the basis of such equivocal findings and this is a matter that is further complicated by the variety of contextual factors in the studies.

Contextual factor: pedagogical interventions. The characteristics of the pedagogical interactions are diverse and indicate many gaps for further examination. The duration of the courses ranges from 2 weeks to 2 years. While the majority of studies examine EE interventions lasting between 3 and 10 months, only one investigates the impact of a short intervention of 2–4 weeks (Huber et al., 2014). Moreover, only two studies look at EE lasting for more than an academic year – Thursby et al. (2009) study a 2-year programme, and Varamäki et al. (2015) followed a cohort through its first 2 years of a Bachelor’s degree course.

Furthermore, when separating the studies into the traditional categories of learning *about*, *for* and *through* entrepreneurship (Jamieson, 1984), it becomes evident that none of the pedagogical interventions can be categorized as learning *about* entrepreneurship. The 17 impact studies are evenly distributed between learning *for* entrepreneurship (nine studies) and *through* it (nine studies),⁴ and no particular differences in terms of positive or negative impact can be observed between these in the SLR sample.

Contextual factor: sample characteristics. Different sample characteristics could have a major impact on how a course is experienced by the participants and the effect of the EE intervention. The educational level of the EE participants is, for instance, a topic for further exploration. An example of is EE impact on primary-school students, as only one study addresses this (Huber et al., 2014). Of the remaining 16 experimental studies, four are about secondary-school students, three concern postgraduate students, and nine examine the impact on undergraduate students. Whether or not a course is compulsory could also have an impact on its effect, and both categories are covered equally in the experimental impact studies.

Bae et al. (2014) show in their meta-analysis that cultural values serve as a moderator of the relationship between EE and entrepreneurial intentions. Hence, the cultural context is another important characteristic that can impact the effect of an EE course. Based on the 17 experimental impact studies, it appears that EE impact studies have predominantly been a Western European exercise (11 studies). There are, however, also three from Africa, two from the USA and one from Australia.

Contextual factor: time frame. In the majority of the 17 experimental impact studies, the post-measurement time is immediately after the completion of the pedagogical intervention. Recent contributions by Volery et al. (2013), Rauch and Hulsink (2015), Gielnik et al. (2015) and Gielnik et al. (2017) have, however, also collected data several months after the intervention. Gielnik et al.

⁴ Varamäki et al. (2015) studies both education *for* and *through* education in the same study.

(2017), for instance, combine measurement right after an EE course with measurements 12 and 28 months after course completion, thereby enabling longitudinal follow-up of development after an EE programme.

DISCUSSION

The findings of this study show that the number of experimental impact studies has increased considerably in recent decades. Nevertheless, 88.3% of the studies can be classified as having a weak pre-experimental design that does not really allow causal claims to be made. This is a major concern in a field that is rapidly expanding and in search of legitimacy among stakeholders such as policymakers, sponsors and educational institutions (Fayolle et al., 2016). In fact, our SLR reveals that only 17 impact studies up to and including 2017 apply a strong experimental design either through true experimental or quasi-experimental design. Hence, there are not that many rigorous studies for policymakers and educators to draw on when making decisions regarding investments and the future development of EE. Obviously, several insightful qualitative studies on outcomes, as well as pre-experimental design studies, provide a valuable understanding of relationships between variables. However, in a fast-moving field in which action and intervention are developing quickly, it is critical that the theory and research needed to justify and explain EE develop simultaneously. Our findings indicate that this has not been the case for strong experimental impact studies on EE. While this is also a challenge for both general and management education (Köhler et al., 2017), the issue is even more pronounced for EE as a young and emerging field. EE scholars are researching new and innovative education initiatives (often with small samples), while established education fields provide more stable conditions to undertake research.

In fact, the qualitative analysis indicates that there is still scant knowledge about the effects of EE as a pedagogical intervention. In general, the majority of the strong experimental impact studies point towards a positive relationship between participation in EE and cognitive, skill-based, affective, conative and behavioural outcomes. However, the SLR also identifies studies that report non-significant and even negative relationships between EE and the impact indicators. The few studies and the small sample sizes of the single studies further complicate the equivocal findings. For example, only four of the 17 studies have a treatment group of more than 200 students. This complicates the application of, for example, meta-analysis, which is a well-recognized approach to summarize effect by combining empirical studies on interventions. For instance, two recent meta-analyses by Bae et al. (2014) and Martin et al. (2013) had to include studies with a weak experimental design in order to draw conclusions. Hence, it is hard to draw categorical conclusions based on the

sample of 17 articles, since their findings appear to point in several different directions, even when the same outcome variables are studied.

Furthermore, with mixed findings, low numbers of experimental studies and small sample sizes, we question whether our findings are valid for other populations in different contexts. EE cannot be treated as a black box, and it is necessary to acknowledge the nuances of EE offered across the world, at different education levels and with quite diverse pedagogics. We agree with Rideout and Gray (2013: 348), who call for a larger pool of methodologically adequate EE studies in order to answer questions such as ‘What type of EE, delivered by whom, within which type of university, is the most effective for this type of student, with this kind of goals, and under these sets of circumstances?’ It is essential to acknowledge the diversity of EE interventions. A compulsory course about entrepreneurship offered to first-year Bachelor’s students in general business could obviously have a different impact on participants from an elective course in an entrepreneurship Master’s in which students start their own companies. There is great variance in EE pedagogics and their impacts will most likely be quite different. By not treating EE as a black box, it will be possible to draw nearer to a more complex understanding of the actual impact of EE interventions.

Thus, the summary of experimental research findings in Table 1 defines important research gaps and points towards future research opportunities. For example, two Spanish impact studies by Sanchez (2011, 2013) concern compulsory courses for secondary and undergraduate students who learn *for* entrepreneurship throughout an 8-month pedagogical intervention. His findings show significant increases in intention, self-efficacy, proactiveness, and risk-taking by EE students. However, when applying Table 1 to identify gaps, there is still much to be explored. Little is known about how Spanish students or those in neighbouring countries will develop during a self-selected elective course or through EE courses for primary education. Furthermore, we do not know anything about either the potential long-term impact, affective outcome measures, or whether EE actually results in entrepreneurial behaviour.

Numerous research gaps could be identified by applying Table 2 in this way. However, we especially want to draw attention to some particularly under-researched issues. For instance, there are no experimental impact studies on courses *about* entrepreneurship. All the identified studies concern learning *for* or *through* entrepreneurship. It is often claimed that learning *about* entrepreneurship does not impact on students, as opposed to the two other approaches (Honig, 2004; Neck and Greene, 2011). However, due to the absence of experimental impact studies on this pedagogical approach there is no robustly researched knowledge to support this view. Moreover, only one study (Huber et al., 2014), from the Netherlands, reports on EE in primary education, which also remains a major research gap. There is also an over-representation of impact studies from Western European countries. Bae et al. (2014) found that the impact of EE is moderated by cultural values and methodologically

rigorous studies from, for example, Eastern Europe or Asia could provide interesting insights into how EE impacts students in other cultural settings.

Accordingly, our findings could serve as an overview of where rigorous EE impact studies are still needed. Furthermore, in line with Nabi et al. (2016a), we find that the majority of impact indicators are short-term, subjective impact measures. As the proof of the pudding is said to be in the eating, there is still major potential for examining long-term impacts such as venture creation and performance. Furthermore, the objective of EE is not necessarily only to increase start-up rates but also to develop the entrepreneurial mindset of students, which can then be used in, for example, existing companies and to enhance students' employability. Thus, novel outcome measures such as intrapreneurial intentions, personal development, social entrepreneurship, employability and career decision making could be fruitful indicators to advance our understanding of EE impact.

The mixed results from impact research also provide an interesting opportunity for further research in order to offer explanations for the equivocal findings. The scenario design by Nabi et al. (2016b) is, for example, an important contribution that sheds light on how the same EE intervention can have different impacts on different students. The suggestion by Von Graevenitz et al. (2010) of a sorting effect, where students become more confident about whether entrepreneurship is a suitable career path for them, also has potential for further exploration. Thus, a decrease in entrepreneurial intentions after EE is not necessarily negative if it is due to enhanced career maturity among participants.

CONCLUSION

The two objectives of this paper are 1) to review the use of experimental research design in EE impact research, and 2) to offer insights into the findings of impact studies that apply a strong experimental design through either true experimental or quasi-experimental design. In doing so, we hope to shed light on the value of applying a strong experimental design in EE impact research and to lay the foundation for a future research agenda. When it comes to the first research objective, the main finding from the study is that there is a substantial lack of strong experimental design in EE impact studies. Of 145 quantitative impact studies identified in ABS-listed ranked journals, only 17 have a true or quasi-experimental design, accounting for 11.7% of the studies. Hence, 88.3% of quantitative impact studies can be characterized as having a weak experimental design. This lack of rigour has severe consequences for the possibility of making inferences and for the generalizability of existing research findings.

Furthermore, with regard to the second research objective, it is evident in the synthesizing of findings from the 17 rigorous impact studies that we still know rather little about the causal link

between EE and entrepreneurial outcome measures. While the majority of impact studies indicate a positive impact, there are also studies with non-significant and even negative impacts on EE outcomes. Hence, based on the findings from the SLR, we call for more true and quasi-experimental studies that can provide robust findings on EE impact. There is a need for more research on the outcome measures identified in the SLR, but there is also potential for exploring novel impact indicators. Intrapreneurship, social entrepreneurship and employability are, for example, outcome measures that remain unexplored in rigorous experimental studies.

An expanding body of rigorous impact studies would also contribute to the development of a more fine-grained understanding of EE and the influence of contextual factors. Context matters in education and EE cannot be treated as a black box. More strong experimental impact studies on the variety of pedagogics, course durations and student samples would accordingly enhance understanding of the nuances of EE impact. As a result, one could get closer to answering important questions such as which pedagogics to apply for a certain group of students if the ambition, for example, is to increase nascent entrepreneurship.

Therefore, although there have been many important research contributions towards an understanding of the complex phenomenon of EE in recent decades, EE impact research has not yet delivered the required empirical findings to EE stakeholders. Teachers and educational institutions need robust evidence on which to base decisions as to when they introduce, execute and develop EE courses. Correspondingly, governments across the world are including EE in educational policies and investing heavily in the implementation of EE. They cannot be expected to continue to do so if EE research does not provide robust evidence of its impact. Hence, the EE research community should take a critical look at the research being conducted and strive to provide EE stakeholders with empirical evidence acquired through methodologically rigorous studies.

Like any methodology, the SLR has its limitations. We acknowledge that the decision to do a journal-led search will deliver different results to those of an open-database search, as would the selection of other search strings. However, by searching impactful journals within EE research, our review highlights a fundamental problem in EE impact research: knowledge about the impact of EE as a pedagogical intervention is scarce. The quality of the research on EE impact is currently lagging behind the thriving development of EE at educational institutions worldwide. As EE continues to spread, it becomes increasingly important for research to justify and explain what is taking place during and after EE courses. For the future, the challenge for EE scholars is to do this with methodologically rigorous studies that can help EE to gain legitimacy both as an educational element and as a research field.

REFERENCES

- Ajzen I (1989) Attitude structure and behavior. In: Pratkanis AR, Breckler SJ and Greenwald AG: *Attitude structure and function*, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 241-274.
- Athayde R (2012) The impact of enterprise education on attitudes to enterprise in young people: an evaluation study. *Education+Training* 54 (8/9): 709-726.
- Audretsch DB, Keilbach MC and Lehmann EE (2006) *Entrepreneurship and economic growth*. New York: Oxford University Press.
- Bae TJ, Qian S, Miao C and Fiet JO (2014) The relationship between entrepreneurship education and entrepreneurial intentions: a meta-analytic review. *Entrepreneurship Theory and Practice*, 38(2): 217-254.
- Blenker P, Elmholdt ST, Frederiksen SH, Korsgaard S and Wagner K (2014) Methods in entrepreneurship education research: a review and integrative framework. *Education+Training*, 56(8/9): 697-715.
- Bloom B, Engelhart MD, Furst EJ, Hill WH and Krathwohl DR (1956) *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. New York: David McKay.
- Campbell DT and Stanley JC (1963) Experimental and quasi-experimental designs for research. *Handbook of research on teaching*. Chicago, IL: Rand McNally.
- Cook TD and Campbell DT (1979) *Quasi-experimentation: Design & analysis issues for field settings*. Boston, MA: Houghton Mifflin.
- DeTienne, DR and Chandler GN (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.
- Elert N, Andersson FW and Wennberg K (2015) The impact of entrepreneurship education in high school on long-term entrepreneurial performance. *Journal of Economic Behavior & Organization*, 111: 209-223.
- Fayolle A, Gailly B and Lassas-Clerc N (2006) Assessing the impact of entrepreneurship education programmes: a new methodology. *Journal of European Industrial Training*, 30(9): 701-720.
- Fayolle A and Liñán F (2014) The future of research on entrepreneurial intentions. *Journal of Business Research*, 67(5): 663-666.
- Fayolle A, Verzat C and Wapshott R (2016) In quest of legitimacy: The theoretical and methodological foundations of entrepreneurship education research. *International Small Business Journal*, 34(7): 895-904.

- Fisher S, Graham M and Compeau M (2008). Starting from scratch: understanding the learning outcomes of undergraduate entrepreneurship education. In: Harrison RT and Leitch C (eds) *Entrepreneurial learning: conceptual frameworks and applications*. New York: Routledge, pp. 313-340.
- Gielnik MM, Uy MA, Funken R and Bischoff KM (2017) Boosting and sustaining passion: A long-term perspective on the effects of entrepreneurship training. *Journal of Business Venturing*, 32(3): 334-353.
- Gielnik MM, Frese M, Kahara-Kawuki A et al. (2015) Action and action-regulation in entrepreneurship: Evaluating a student training for promoting entrepreneurship. *Academy of Management Learning & Education*, 14(1): 69-94.
- Honig B (2004) Entrepreneurship education: Toward a model of contingency-based business planning. *Academy of Management Learning & Education*, 3(3), 258-273.
- Huber LR, Sloof R and Van Praag M (2014) The effect of early entrepreneurship education: evidence from a field experiment. *European Economic Review*, 72: 76-97.
- Jamieson, I (1984) Schools and enterprise. In: Watts AG and Moran P (eds): *Education for enterprise*. Cambridge, UK: CRAC, pp. 19-27.
- Johnson B and Christensen L (2012) *Educational research - quantitative, qualitative, and mixed approaches*. London: Sage Publications.
- Kamovich U and Foss L (2017) In search of alignment: a review of impact studies in entrepreneurship education. *Education Research International*, 2017 (1): 1–15
- Karlsson T and 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.
- Kolvereid L and Moen Ø (1997) Entrepreneurship among business graduates: does a major in entrepreneurship make a difference? *Journal of European Industrial Training*, 21(4): 154-160.
- Köhler T, Landis RS and Cortina JM (2017) From the editors: establishing methodological rigor in quantitative management learning and education research. *Academy of Management Learning & Education*, 16(2): 173-192.
- Kozlinska I (2016). *Evaluation of the outcomes of entrepreneurship education revisited evidence from Estonia and Latvia*. PhD Thesis, University of Tartu, Tartu, Finland.
- Kraiger K, Ford JK and Salas E (1993) Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, 78(2): 311-328.

- Kyrö P (2008) A theoretical framework for teaching and learning entrepreneurship. *International Journal of Business and Globalisation*, 2(1): 39-55.
- Lange J, Marram E, Jawahar AS, Yong W and Bygrave W (2011) Does an entrepreneurship education have lasting value? *Frontiers of Entrepreneurship Research*, 31(6): 209-224
- Lorz M, Mueller S and Volery T (2013) Entrepreneurship education: a systematic review of the methods in impact studies. *Journal of Enterprising Culture*, 21(2): 123-151.
- Martin BC, McNally JJ and Kay MJ (2013) Examining the formation of human capital in entrepreneurship: A meta-analysis of entrepreneurship education outcomes. *Journal of Business Venturing*, 28(2): 211-224.
- Martinez AC, Levie J, Kelley DJ, Sæmundson RJ and Schøtt T (2010) Global Entrepreneurship Monitor special report: a global perspective on entrepreneurship education and training. Report, Global Entrepreneurship Monitor, Boston, United States.
- Mentoor ER and Friedrich C (2007) Is entrepreneurial education at South African universities successful? An empirical example. *Industry and Higher Education*, 21(3): 221-232.
- Mertens D (2010) *Research and evaluation in education and psychology*. London: Sage Publications
- Mets T, Kozlinska I and Raudsaar M (2017) Patterns in entrepreneurial competences as the perceived learning outcomes of entrepreneurship education: The case of Estonian HEIs. *Industry and Higher Education*, 31(1): 23-33.
- Mwasalwiba ES (2010) Entrepreneurship education: a review of its objectives, teaching methods, and impact indicators. *Education+Training*, 52(1): 20-47.
- Nabi G, Liñan F, Fayolle A, Krueger NF and Walmsley A (2016a) The impact of entrepreneurship education in higher education - a systematic review and research agenda. *Academy of Management Learning & Education*, 16(2): 277-299.
- Nabi G, Walmsley A, Liñan F, Akhtar I and Neame C (2016b) Does entrepreneurship education in the first year of higher education develop entrepreneurial intentions? The role of learning and inspiration. *Studies in Higher Education*, 43(3): 452-467.
- Neck HM and Greene PG (2011) Entrepreneurship education - known worlds and new frontiers. *Journal of Small Business Management*, 49(1): 55-70.
- O'Connor A (2013) A conceptual framework for entrepreneurship education policy. *Journal of Business Venturing*, 28(4): 546-563.
- Oosterbeek H, Van Praag M and Ijsselstein A (2010) The impact of entrepreneurship education on entrepreneurship skills and motivation. *European Economic Review*, 54(3), 442-454.
- Peterman NR and Kennedy J (2003) Enterprise education: influencing students' perceptions of entrepreneurship. *Entrepreneurship Theory & Practice*, 28(2): 129-144

- Pittaway L and Cope J (2007) Entrepreneurship education: a systematic review of the evidence. *International Small Business Journal*, 25(5): 479-510.
- Rauch A and 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.
- Rideout EC and Gray DO (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.
- Rynes SL and Brown KG (2011) Where are we in the 'long march to legitimacy?' Assessing scholarship in management learning and education. *Academy of Management Learning & Education*, 10(4), 561-582.
- Sanchez JC (2011) University training for entrepreneurial competencies: Its impact on intention of venture creation. *International Entrepreneurship and Management Journal*, 7(2): 239-254.
- Sanchez JC (2013) The impact of an entrepreneurship education program on entrepreneurial competencies and intention. *Journal of Small Business Management*, 51(3): 447-465.
- Shadish W, Cook TD and Campbell DT (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Slavin RE (2002) Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31(7): 15-21.
- Souitaris V, Zerbinati S and Al-Laham A (2007) Do entrepreneurship programmes raise entrepreneurial intention of science and engineering students? *Journal of Business Venturing*, 22(4): 566-591.
- Tranfield D, Denyer D and Smart P (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3): 207-222.
- Thursby MC, Fuller AW and Thursby J (2009) An integrated approach to educating professionals for careers in innovation. *Academy of Management Learning & Education*, 8(3): 389-405.
- Varamäki E, Joensuu S, Tornikoski E and Viljamaa A (2015). The development of entrepreneurial potential among higher education students. *Journal of Small Business and Enterprise Development*, 22(3): 563-589.
- Volery T, Müller S, Oser F, Naepflin C and Rey N (2013) The impact of entrepreneurship education on human capital at upper-secondary level. *Journal of Small Business Management*, 51(3): 429-446.
- Von Graevenitz G, Harhoff D and Weber R (2010) The effects of entrepreneurship education. *Journal of Economic Behavior & Organization*, 76(1): 90-112.

Wang CL and Chugh H (2014) Entrepreneurial learning: past research and future challenges.

International Journal of Management Reviews, 16(1): 24-61.

Wilson F, Kickul J and Marline D (2007) Gender, entrepreneurial self-efficacy, and entrepreneurial career intentions - implications for entrepreneurship education. *Entrepreneurship Theory and Practice*, 31(3): 387-406.

Table 2. Overview of the 17 rigorous experimental impact studies on entrepreneurship education.

Authors	Year	Research design				Outcome measure												
		TED	QED	Cognitive		Skill-based			Affective		Conative			Behavioural				
				Knowledge	Traits	Skills	Passion/ inspiration	Ent. attitude	Subjective norm	Ent. intention	Feasibility	Nascency	Venture creation					
Gielnik et al.	2017	x							+					+			+	
Nabi et al.	2016		x		+									n.s.				+
Gielnik et al.	2015	x			+									+				+
Rauch and Hulsink	2015		x										+	+				+
Varamäki, Joensuu, Tomikoski and Viljamaa	2015		x										n.s./-	n.s./-				n.s.
Huber, Sloof and van Praag	2014	x			n.s.	+		+										+
Sanchez	2013																	+
Volery et al.	2013		x		+			+										+
Karlsson & Moberg	2013		x		n.s.	n.s./+												n.s.
Athayde	2012		x															+
Sanchez	2011		x															+
Osterbeek, van Praag and Ijsselstein	2010		x					n.s.	n.s./+									+
Thursby, Fuller and Thursby	2009		x							+								+
Mentoor and Friedrich	2007		x					n.s./-										+
Souitaris, Zabinati and Al-Laham	2007		x											n.s.	+			n.s.
De Tienne and Chandler	2004			x														+
Peterman and Kennedy	2003			x														+

Authors	Course characteristics							Education level			Compulsory			
	Country	Course duration	Time of post measurement	Final sample size	About	For	Through	Primary	Secondary	Undergrad.	Postgrad.	Yes	No	Unknown
Gielenik et al.	Kenya	12 weeks	T2: right after T3: 12 months after T4: 28 months after	125 treatment 102 control			x			x				x
Nabi et al.	UK	12 months	T2: right after	89 treatment 61 control		x				x				x
Gielenik et al.	Uganda	12 weeks	T2: 1 month T3: 9 months after T2: 6 months	162 treatment 142 control			x			x				x
Rauch and Hulsink	Netherlands	1 year	T2: 6 months	62 treatment 18 control		x				x				x
Varamäki, Joensuu, Tomikoski and Viljamaa	Finland	Several courses	T1: 1st year T2: 2nd year	66 treatment 131 control		x				x				x
Huber, Sloof and van Praag	Netherlands	2-4 weeks	T2: right after	1727 treatment 684 control			x							x
Sanchez	Spain	8 months	T2: right after	347 treatment 363 control		x				x				x
Volery et al.	Switzerland	6-12 months	T2: right after T3: 4-5 months after T2: right after	291 treatment 102 control 51 treatment 21 control			x			x				x
Karlsson and Moberg	Denmark	9 months	T2: right after	200 treatment 76 control			x			x				x
Athayde	London, UK	9 months	T2: right after	200 treatment 76 control			x			x				x
Sanchez	Spain	8 months	T2: right after	404 treatment 460 control		x				x				x
Osterbeek, van Praag and Ijsselstein	Netherlands	2 semesters	T2: right after	104 treatment 146 control			x			x				x
Thursby, Fuller and Thursby	US	2 years	T2: right after	71 treatment 15 control post		x				x				x
Mentoor and Friedrich	South Africa	1 semester	T2: right after	418 treatment 45 control		x				x				x
Souitaris, Zabinati and Al-Laham	France and UK	5 months	T2: right after	124 treatment 126 control		x				x				x
DeTienne and Chandler	US	1 semester	T2: right after	71 treatment 59 control		x				x				x
Peterman and Kennedy	Australia	5 months	T2: right after	109 treatment 111 control			x			x				x

APPENDIX I – READING GUIDE

Reading guide								
1. General information								
1a. Author(s)								
1b. Year of publication								
1c. Article title								
1d. Journal								
2. Theoretical positioning								
2a. Theoretical framework								
3. Impact								
3a. Impact measures	Cognitive Knowl. <input type="checkbox"/>	Cognitive Traits <input type="checkbox"/>	Skill-based Skills <input type="checkbox"/>	Affective Passion Inspo. <input type="checkbox"/>	Affective Ent. att. <input type="checkbox"/>	Affective Sub. Norm <input type="checkbox"/>	Conative Ent. int. <input type="checkbox"/>	Conative Feasib. <input type="checkbox"/>
	Behavioral Nascency <input type="checkbox"/>	Behavioral Start-up <input type="checkbox"/>						
3b. Measurement items								
3c. Reported impact								
3d. Control variables (CV)								
3e. Reported CV effect								
4. Methodology								
4a. Research design	TED: RPPC <input type="checkbox"/>	TED: RS4 <input type="checkbox"/>	QED: NEPPC <input type="checkbox"/>	PED: 1PO <input type="checkbox"/>	PED: NEPO <input type="checkbox"/>	PED: 1PP <input type="checkbox"/>		
4b. Data collection method	Questionnaire <input type="checkbox"/>					Other:		
4c. Follow-up length								
5. Sample characteristics								
5a. Sample size								
5b. Education field	Business <input type="checkbox"/>	Science <input type="checkbox"/>	Humanities <input type="checkbox"/>	Social <input type="checkbox"/>	Health <input type="checkbox"/>	Education <input type="checkbox"/>		
	Primary <input type="checkbox"/>	Secondary <input type="checkbox"/>	Tertiary <input type="checkbox"/>					
5c. Education level								
5d. Country								
6. Intervention characteristics								
6a. Course description	About <input type="checkbox"/>	For <input type="checkbox"/>	Through <input type="checkbox"/>	JA-YE <input type="checkbox"/>				Other
	Yes <input type="checkbox"/>	No <input type="checkbox"/>						
6b. Compulsory								
6c. Duration								
6d. Total work hours								
6e. Team work	Yes <input type="checkbox"/>	No <input type="checkbox"/>						
7. Analysis								
7a. Data analysis method								
7b. Key findings								