social sciences

MDPI

Project Report

Co-Creative Action Research Experiments—A Careful Method for Causal Inference and Societal Impact

Arjen van Witteloostuijn [1,2,3], **Nele Cannaerts** [4], **Wim Coreynen** [2,5], **Zainab Noor el Hejazi** [3], **Joeri van Hugten** [1,*], **Ellen Loots** [6], **Hendrik Slabbinck** [7] and **Johanna Vanderstraeten** [3]

1   School of Business and Economics, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands; a.van.witteloostuijn@vu.nl
2   Antwerp Management School, University of Antwerp, 2000 Antwerp, Belgium
3   Department of Management, Faculty of Business and Economics, University of Antwerp, 2000 Antwerp, Belgium; zainabnoor.elhejazi@uantwerpen.be (Z.N.e.H.); johanna.vanderstraeten@uantwerpen.be (J.V.)
4   Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam, 3000 DR Rotterdam, The Netherlands; cannaerts@essb.eur.nl
5   School of Management, Zhejiang University, Hangzhou 310058, China; wimcoreynen@zju.edu.cn
6   Arts and Culture Studies Department, Erasmus School of History, Culture and Communication, Erasmus University, 3062 PA Rotterdam, The Netherlands; loots@eshcc.eur.nl
7   Department of Marketing, Innovation and Organisation, Ghent University, 9000 Gent, Belgium; hendrik.slabbinck@ugent.be
*   Correspondence: j.g.w.j.van.hugten@vu.nl

check for updates

Abstract: The rigor-versus-relevance debate in the world of academia is, by now, an old-time classic that does not seem to go away so easily. The grassroots movement Responsible Research in Business and Management, for instance, is a very active and prominent advocate of the need to change current research practices in the management domain, broadly defined. One of its main critiques is that current research practices are not apt to address day-to-day management challenges, nor do they allow such management challenges to feed into academic research. In this paper, we address this issue, and present a research design, referred to as CARE, that is aimed at building a bridge from rigor to relevance, and vice versa. In so doing, we offer a template for conducting rigorous research with immediate impact, contributing to solving issues that businesses are struggling with through a design that facilitates causal inference.

Keywords: rigor; relevance; research design

1. Introduction

Management, broadly defined, is an academic discipline that is deeply rooted in practice. Originally, the discipline of management was established in a university environment in order to feed into teaching programs meant to educate the managers of the future by contributing to the introduction of evidence-based managerial practices (Khurana 2010). This institutional format developed into business schools, in which academic scholars joined forces with practice. Ever since, an ongoing debate has been evolving, and still does so, around this very rigor-versus-relevance tension (see, e.g., Gulati et al. 2007). On the one hand, as one argument goes, management as a discipline is so much focused on practical relevance that scholarly rigor is being sacrificed along the way. On the other hand, another—opposite—argument is that management as an academic discipline that has started to adopt scholarly practices, which comes at the expense of relevance. So, many business schools host academics

that engage in an inward-oriented "l'art pour l'art" game that no longer resonates in practice, obsessed by theoretical fetishism that is highly self-referential in nature (Birkinshaw et al. 2014).

In the current paper, we do not directly speak to this debate in the sense of arguing one way or the other. Rather, we start from the observation that, by and large, the state of the art in management is neither sufficiently rigorous, nor relevant enough. We thus build on the argument of the Responsible Research in Business and Management (RRBM) grassroots movement (cf. Tsui 2013a, 2013b) that argues that much academic management research might be rigorous, but fails to influence practice. Specifically, our argument is two-fold. First, fundamentally, management research is still not rigorous enough by not being good at causal inference. That is, notwithstanding the application of advanced econometric techniques, by far the majority of studies in management are essentially correlational in nature. This is further complicated by the statistical hocus pocus that is oftentimes incorrectly executed and meant to suggest causal relations (Antonakis et al. 2010). However, to be relevant, either from a scholarly or a practical point of view, the identification of causality is key (e.g., Gow et al. 2016). Second, and equally fundamentally, even if extant research would be sufficiently causal in nature, real relevance is hard to find. Indeed, as argued by proponents of RRBM, management scholars tend to play the academic "high impact factor game" (cf. van Witteloostuijn 2016), not really caring about practical relevance at all. All these obligatory paragraphs on "managerial implications" in all these scholarly publications only pay lip service to practice, as the number of real-world practitioners reading, let alone applying insights from, all these top or not-so-top academic journals is close to zero.

The above relates to a broader critique, that the social sciences, of which management is just one branch, are insufficiently solution-oriented (Watts 2017). In the present paper, our key argument is that this implies that we, as a scholarly community (in management and beyond), should aim for research that is both more rigorous and more relevant—so, no 'either-or' here, but 'and'. Of course, we are not the first to argue this; and of course, we are aware of initiatives complementary to ours (see, e.g., Gray and Purdy 2018). However, we argue in favor of a very specific research design that we believe can be particularly powerful to close the rigor vis-à-vis relevance gap, by strengthening both sides of the divide. We refer to this design as Co-creative Action Research Experiments, or CARE. Below, we will not only explain what CARE is all about in theory, but we will also present a concrete example of a CARE-based project.

In the next section, we first explain that the solution is not so much in applying even more fancy econometric techniques or developing even more incomprehensible theories, nor in adopting different epistemologies, but rather to dare to develop rigorous causal research designs in co-creation with practice. Key here is to combine relevance (hence co-creative action) with causal inference (hence experimental research). This, we argue, requires a CAREful design in the form of Co-creative Action Research Experiments. In the section after that, an (ongoing) example of such a design, one with which we have been experimenting since 2016, is introduced. Subsequently, we briefly present an example of an analysis on the basis of CARE-collected data. In the final section, we summarize our plea to change current research practices in management (and the social sciences, more broadly) in an attempt to provide the rigor and relevance needed to benefit both academia and practice.

2. Toward a Co-creative Action Research Experimental Design

We cannot critique management for not producing a sufficient number of studies—quite the contrary. Over the post-war decades, the output of scholarly management work has exploded, and this exponential growth trend is unlikely to come to an end any time soon. With a country such as China entering into the front-end of the academic output machinery, this growth is actually accelerating even more. So, rather than underproduction, overproduction might be an issue. However, the critique that this massive stock and flow of research are disconnected from real-world managerial practice is anything but silenced. Indeed, as argued elsewhere (e.g., Starbuck 2016; van Witteloostuijn 2016), like many other scientific disciplines, management has evolved into an inward-oriented scholarly community with incentives to engage in unproductive, incorrect and questionable research practices.

This paper is not the place to reiterate this diagnosis of the current state of affairs (see, e.g., Meyer et al. 2017). Suffice to say that we fail to really accumulate knowledge by turning to *p*-hacking and HARKing, and by not engaging in replication (to mention just a few, albeit critical, malpractices).

The result is that we, over many decades, have created an academic management community that is self-referential. The first bell that starts ringing when management scholars refer to 'impact' is not the common-sense one related to improving practice, but that associated with the impact factor of our scholarly journals, the accumulated number of citations, and the *h*-index. This is not to say that all our research lacks practical relevance, and that all our teaching involves an academic fantasy. But what this does imply is that our primary motive is to produce scholarly output for these high-impact journals to boost our academic reputation (and career, for that matter). This is at the root of the imbalance so convincingly and forcefully pointed at by the RRBM movement, and many likeminded critiques (e.g., Tsui 2013a, 2013b). Most academic management research, by far, is not aimed to impact practice at all. Our research agenda is not primarily driven by the needs of practice, but by what is looked for by our (top) academic journals, with any overlap being accidental (Walker et al. 2019). This explains, in the large part, why academic research is not influencing managerial practice that much, if at all. Rather, managerial practice by far outpaces academic research in terms of innovative practices, both organizational and strategic (Khurana 2010).

Disappointingly, much academic research is rigorous in a particular way, and not one inducive to producing any solid evidence base for managerial practice. The latter requires rigor in causal inference, which is rare in the management domain (Gow et al. 2016; Maula and Stam 2019). Would such evidence emerge, this often can be seen as a case of 'collateral benefit'. To impress readers and reviewers, much of the extant (quantitative) work excels in advanced multivariate statistics, following the fashion of the day. However, much of this advancement is only trying to come as close as possible to causal identification, but without really getting there, due to the very nature of the data associated with the research design that is applied. Panel analyses com closest, with an *n* and *t* large enough to introduce a meaningful lag structure in the specification in combination with theoretically plausible and econometrically valid instrumentation, approximating what can be referred to as a 'natural experiment' (Reeb et al. 2012). Regrettably, by far the majority of the extant work in management does not even come close to this ideal, either involving cross-sectional data or a too small *n* and/or *t*, or not introducing a (credible) instrument (Aguinis and Edwards 2014; Bergh et al. 2017; Maula and Stam 2019).

One rather popular way out is argued to be processual case studies (see, e.g., Dawson 2019). More generally, many argue that qualitative work in the form of any of the many variants of a rich case study design, from longitudinal (comparative) case analysis to qualitative comparative analysis (known as QCA), provides the toolkit for causal inference (Fiss 2011; Van Burg et al. 2020). Here, we do not engage in questioning the overall validity of this claim. This may or may not be true. In any case, what such work cannot provide is (a) a rigorous estimate of effect sizes, (b) systematic control for alternative explanations, and (c) generalization over large *n*s (preferably, drawn from different populations). Moreover, by far the majority of qualitative work involves the analysis of historical material (as do much, if not close to all, quantitative studies), implying a difficult-to-avoid hindsight bias. All this makes the application of qualitative designs, however valuable for all kinds of other reasons (such as theory development or processual insight), insufficient to accumulate rigorous causal evidence for informing managerial practices.

Our central claim here is that in order to solidly bridge the rigor-relevance gap, we need to turn to a research design developed precisely to engage in causal identification in a context seen as relevant by practice. This implies three essential attributes. First, such a research design has to be rigorous in the sense of reliably and validly offering the opportunity to engage in causal inference (Van de Ven and Johnson 2006). As we will argue below, this implies a design that is a first-best equivalent to or a second-best approximation of a field experimental set-up, coming close to the randomized control trial (RCT) ideal in economics and medicine. For understandable reasons, RCT field experiments are rarely within reach within management; for the wrong reasons, lab experiments

(RCT or otherwise) have a hard time in management (van Witteloostuijn 2015). This aspect of our design involves the rigor-aka-causal inference side of the design template we will detail below. Second, to safeguard relevance, close collaboration with practice is recommended. In a process of co-creation, both parties can develop research questions and can identify field settings that are of immediate benefit to and/or relevant for managerial practice (Reason 2006; Leitch 2007; Kieser and Leiner 2009). This is the relevance-aka-co-creation action research side of our design template. Third, multidisciplinary research is needed to understand the ever-increasing complexity of our surrounding world (e.g., Nopens et al. 2019). This is exactly what we will be doing, by combining insights from disciplines such economics, psychology and sociology (with sub-domains such as human resource management, international business, entrepreneurship, organizational behavior, and strategic management) in the design template we suggest in the current paper.

Our Co-creative Action Research Experiments (i.e., CARE) design involves six essential elements, which we briefly introduce below. Central to the CARE design is action in the form of intervention. Here, an intervention is defined as a 'treatment' of an entity, similar to that in the RCT context. For instance, such an intervention may involve a leadership training tailored at entrepreneurs of small- and medium-sized enterprises (SMEs). In all, CARE involves six key elements:

(a) **Co-creation**: A sine qua non is that practice is involved early on. Together with representatives of the research population (say, public bureaucracies, multinational enterprises or small businesses), the research team identifies the central question(s), and takes this (these) as the steppingstone(s) for co-developing the full cycle associated with the research design. For instance, representatives from practice can and should contribute to selecting the key (dependent, independent and control) variables, gaining access to the field, developing intervention strategies, and disseminating key insights. Such an involvement can lead to overcoming a major hurdle in collaborative action research (Kieser and Leiner 2009, p. 528; Van de Ven and Johnson 2006).

(b) **Qualitative information**: Context is key. Whatever the focus of study, and that of the action or intervention, each individual, group or organization is different, featuring specific idiosyncrasies. Such idiosyncrasies are hard, if at all, to capture through the usual control variables strategy. The latter is too crude a sieve, always being associated with an omitted variables bias. Hence, by adopting mixed methods, qualitative information is collected that can facilitate putting a specific 'unit of intervention' in its specific perspective (Reason 2006; Van Burg et al. 2020).

(c) **Quantitative measurement**: Throughout the intervention and action cycle, to the extent feasible and possible, the essential (dependent, independent and control) variables have to be measured quantitatively. The source of quantitative data is a mixture tailored to what has to be measured. Objective data (say, financial figures from annual reports) are combined with subjective measures (e.g., through questionnaire scales). Subjective data can tap into the respondents' controlled (i.e., survey items) or automatic (i.e., implicit tests) responses. It is critical to have pre and post-intervention measures, as well as information regarding key features of the intervention (e.g., DeTienne and Chandler 2004).

(d) **Action guidance**: An essential ingredient of the design is the active involvement of practice. Apart from the element of co-creation introduced above, practice is actively engaged in guiding the intervention action, as well as providing access to the field in which to experiment and measure. In this way, the research has immediate impact, reciprocal learning is facilitated (from academia to practice, and vice versa), and dissemination of evidence is accommodated (directly in the short run, and indirectly in the long run) (Reason 2006; Leitch 2007).

(e) **Experimental intervention**: Causal inference requires an experimental intervention. Ceteris paribus, any post-intervention change can be causally attributed to the intervention. Of course, this ceteris paribus clause is key. In the ideal world, this involves a randomized controlled trial. However, in the field, noise is inevitable, and random results might wrongly be interpreted as actual results. This is why control variables still have to be included, why qualitative information

must provide background, why dialogue with practice has to be repeatedly organized, and why a control group analysis is needed (Prowse and Camfield 2013).

(f)　**Matched-pair follow-up**: In the field of management, a randomized controlled design tends to be out of reach, for a variety of reasons, notably ethical and practical considerations. Hence, the within-subject design implied by the intervention strategy, as introduced above, is combined with a between-subject analysis by a matched-pair follow-up. That is, each intervention subject (this can be anything, from individuals or groups to organizational units or full-blown enterprises) is matched with a non-intervention twin (e.g., Campbell 2013). The key is, of course, the selection of matching criteria (say, size, sector and profitability in the case of small businesses) and access to sufficient information about the non-intervention twin to run meaningful matched-pair analyses.

Below, we bring the above six essential elements of CARE to life in a real-life coaching example, which we refer to as Ambition in Entrepreneurship (*AiE*, in Dutch, *Ambitie in ondernemen*, or *AiO*). Please note that the research design of the *AiE* project can be applied to any other project attributing great importance to the rigor-relevance bridge. We use *AiE* as an example to illustrate how a CARE design could and should look—nothing more and nothing less.

3. Ambition in Entrepreneurship (*AiE*)

Background

In 2015, the first author, being affiliated to the two founding academic institutions of the *AiE* project (i.e., the Antwerp Management School (AMS, Belgium) and the Faculty of Business and Economics of the University of Antwerp (UAntwerp, Belgium)), joined forces with UNIZO (*Unie van Zelfstandige Ondernemers*), which is the Flemish association for small and medium-sized enterprises (SMEs), to draft a proposal to be submitted to the Flemish Agency for Innovation and Entrepreneurship (VLAIO—*Vlaams Agentschap voor Innovatie en Ondernemerschap*). VLAIO had launched a research call asking for large-scale proposals of consortia with the aim to boost Flemish entrepreneurship. The proposal was granted, with the project running from May 2016 to May 2020 (a follow-up project in the form of an updated version of *AiE* will run from July 2020 to July 2024, with about half of this paper's author team being involved).

Taking co-creation seriously, the research design was developed during an iterative process by a founding team with members from both academia (i.e., AMS and UAntwerp) and practice (i.e., UNIZO). This research design was then applied to a coaching service for ambitious entrepreneurs: *AiE*. One key attribute of the development of both the overall research design (see Figure 1) and the *AiE* coaching service is worth emphasizing. As said, co-creation is essential, involving close collaboration and intense dialogue between academia and practice, in order to keep working on constructing and maintaining a solid rigor-relevance bridge. Apart from the fine-tuning of measures, interpretations, and processes over time of the *AiE* coaching service, often after extensive discussion between practice and academia, two defining phases are worth explaining. The first relates to the development of the overall research design and its *AiE* application, and the second to the fine-tuning of the actual coaching service of the *AiE* project.

First, before the first entrepreneur was enrolled in the coaching service, we took about six months to carefully develop the full research design. Doing so, the co-creation team took an agile development approach, under the guidance of a professional consultant. In a series of meetings, the overall research design was developed, after which it was applied to the *AiE* coaching service. Thus, during these meetings, the full research design was developed (from recruitment and intake to advice and follow-up), and adequate measures for the actual *AiE* coaching service were selected. All concepts were included for two essential reasons: (i) practical relevance; and (ii) scholarly evidence. Second, midway through the project, the academic team decided to carefully and systematically evaluate the actual measures of the *AiE* coaching service, again in close collaboration with the representatives from practice. As a result, a few measures were shortened, and a few were replaced by new ones, either because the initial

measures turned out to be problematic in practice, or because we sought to further broaden the scope of our theoretical reach. Note that this midway redesign was rather limited, leaving the processual set-up, and thus also the overall research design, fully intact, only involving about ten percent changes in the measurement toolkit. In the next section, we will indicate the measures that were removed, shortened or added when we discuss the associated measurement tool of the *AiE* coaching service.
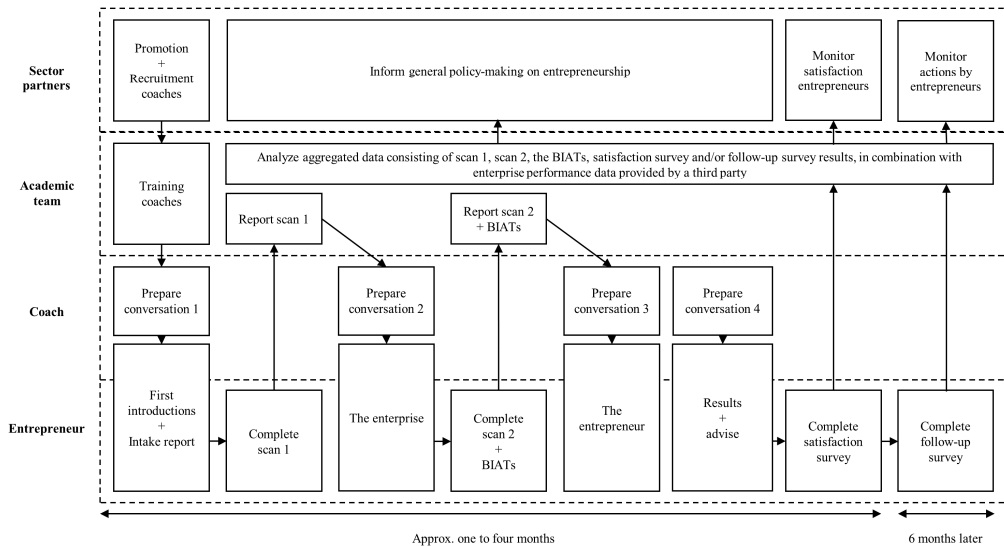

Figure 1. Research design and application to the *AiE* coaching service.

Figure 1, fully introduced and explained in the next section, provides an overview of all the research design's final elements, including the timeline and interdependencies of the design one could follow during, e.g., a coaching trajectory in close collaboration with academia, like we did with the *AiE* coaching service. The whole research cycle, including measurement instruments and intervention strategies, was initially co-developed by the founding team of AMS/UAntwerp and UNIZO. In the course of time, new members of the project team, from both academia and practice, became involved and contributed to changing and improving the initial design (for instance, a mid-way re-design of the quantitative measurement instruments). Apart from AMS/UAntwerp and UNIZO, two additional co-creation partners are worth mentioning.

The first is Graydon, a credit-counselling services provider. Graydon delivers the longitudinal secondary (demographic and financial) information that is appended to the primary qualitative and quantitative data. The second is an expanding group of SME coaches, in charge of guiding the intervention as executed in and with SMEs. They thus guide the SMEs through our research design, and apply it in practice during our *AiE* project (again, see Figure 1 for the different steps of the research design). To facilitate the dialogue with coaches, all coaches were trained before entering into the intervention arena, after which they became full-blown members of what we referred to as the 'learning network'. This learning network is the group of involved people (scholars, representatives from UNIZO, a policymaker from VLAIO, Graydon, and all coaches) of about 25 that met, and still meet, on about a bi-monthly basis.

Content-wise, the *AiE* coaching service to which we apply our research design involves an evidence-based coaching service for ambitious entrepreneurs (Hermans et al. 2014), with the fee for this service being relatively low due to the VLAIO subsidy. This is why our project is referred to as 'Ambition in Entrepreneurship' (*AiE*). The aim is to support entrepreneurs in their attempt to reach their ambition. In principle, this ambition can be anything, from high growth or healthy profitability to innovation or internationalization. In practice, the majority of the participating entrepreneurs seek support to increase their growth and/or profitability. This is perfectly aligned with the Flemish government's ambition to stimulate growth among Flemish SMEs, as reflected in the larger VLAIO

program of which our project is part. Indeed, much of the extant work on high-growth entrepreneurship has revealed, time and again, that sustainable, 'gazelle-like', high growth is the rare exception rather than the rule among SMEs (see, e.g., Coad 2009). As we will detail below, the *AiE* project takes a comprehensive—multi-disciplinary and multi-level—perspective in order to provide evidence-based diagnosis and advice to ambitious entrepreneurs, including an extensive coaching trajectory and follow-up process.

Before introducing all elements of our research design and its application to the *AiE* coaching service, one disclaimer is in place. Our comprehensive data-collection effort involves dozens of concepts and measures. Introducing all these in detail, including underlying theories, would require a publication of book length, even more so if we would add a review of the associated literatures that spell out theory or evidence regarding the immense number of possible relationships—direct and indirect, mediation and moderation, and linear and non-linear. Hence, as this paper is about research design, and not about any specific theoretical literature or empirical study, we only list all our concepts and measures with references to studies where the interested reader can find (much) greater empirical, psychometric, and theoretical detail.

4. The Research Design and its *AiE* Application

All in all, four parties are involved during the overall research design and its application to the *AiE* coaching service: an academic partner (in our case, AMS/UAntwerp), a business association and its coaches (in our case, UNIZO and its coaches), a credit-counselling services provider of secondary objective data (in our case, Graydon), and the clients/participants/respondents (in our case, entrepreneurs and their SMEs). UNIZO is responsible for attracting ambitious entrepreneurs and professional coaches. We—as the academic partner—invest(ed) heavily in training all involved coaches to become experts in our trajectory, which follows the research design visualized in Figure 1. For instance, we drafted background material and organized lectures (recorded and put online) to carefully and systematically explain all theoretical concepts and empirical measures (and the underlying theories) included in the project. Additionally, a new coach was mentored by an experienced colleague, and first assisted an experienced coach in one or two coaching trajectories before performing this service independently. Moreover, during the long chain of bi-monthly learning network meetings, we further explained concepts and measures, exchanged experiences by discussing real-world cases, and presented results from analyses of the data. An important output of this ongoing dialogue is a series of guidelines and templates in which we spell out how to interpret the data, and how to translate this information into concrete advice by identifying personal and strategic mismatches (see below).

When turning our attention to the clients and their coaches, we intentionally opted for a prolonged period (from one to four months) of intensive data gathering with the participating entrepreneurs and the trained coaches, followed by a regular follow-up to trace the entrepreneurs' progress (or lack thereof). Although this might increase the social desirability bias (for which we can control) (Toh et al. 2006), this does decrease the likelihood of common-method variance bias to (close to) zero (Chang et al. 2010) and provides the opportunity to follow-up on the actual impact, if any, of the coaching program. The latter is not only done during and after the coaching trajectory, but also through a matched-pair analysis, where we compare each SME that went through the coaching trajectory with a 'twin' not participating in our program.

All in all, our research design consists of eight overarching steps, each requiring intensive engagement of several partners involved and taking account of all six CARE elements. In a nutshell, we visualize the program in Figure 1. Below, in **bold italics**, we indicate the output that is produced along the way. In Box 1, we provide an overview of this output, referred to as the ***AiE Toolkit***.

Box 1. Ambition in Entrepreneurship Toolkit.

Online Lecture: In an online lecture (in Dutch), the whole design is explained by the academic team, including all the constructs and measures.

Enterprise Manual: In this manual, all constructs and measures included in the *Enterprise Scan* are defined and explained.

Entrepreneur Manual: In this manual, all constructs and measures included in the *Entrepreneur Scan* and *BIATs* (Brief Implicit Association Tests) are defined and explained.

Trade Report: This is a document with all demographic, financial and other information that can be automatically uploaded through Graydon.

Intake Report: This report includes semi-structured qualitative information about the SME and the entrepreneurs.

Enterprise Scan: This is a survey with dozens of items to measure dozens of firm- and environmental-level constructs (see Table 1).

Enterprise Report: This report provides and explains all scores from the *Enterprise Scan*.

Entrepreneur Scan: This is a survey with dozens of items to measure dozens of attitudes and attributes of the individual entrepreneur (see Table 2).

BIATs: This document provides all rankings regarding the three BIATs, and explains their interpretation.

Entrepreneur Report: This report provides and explains all scores from the *Entrepreneur Scan* and *BIATs*.

Advisory Report: This report offers advice to the entrepreneur and her or his enterprise, focusing on ten or so issues that stand out.

Mirror Report: This report identifies and explains the ten most prominent misfits across all the scans and BIATs.

Satisfaction Survey: This survey is administered immediately after the trajectory to assess the entrepreneur's satisfaction with the coaching service.

Follow-Up Survey: This survey is administered about six months after the trajectory to assess the extent to which the entrepreneur implemented the advice, and the subjective evaluation of the (future) impact.

Stand-Alone Panel: All information collected throughout and after the trajectory is pooled in this panel dataset.

Twin Panel: This panel includes Graydon information available for all twins of the SMEs that participated in the program.

---

5. Promotion, Recruitment and Training

To be able to target ambitious entrepreneurs, we promote the coaching service to local entrepreneurs through the UNIZO network and AMS's communication channels (Step #1 in Figure 1). Initially, our target sample consisted of entrepreneurs currently in a transition phase. Transition phases follow the traditional company life cycle path, distinguishing five key transitions (Lester et al. 2003): (1) self-employed entrepreneurs about to hire their first employee; (2) micro-firms growing into small firms; (3) small firms growing into medium-sized enterprises; (4) firms that are in decline or have recently experienced setbacks; (5) and firms that are about to stop, or are looking to be acquired. Quickly into the project, in close alignment with VLAIO, the primary target was defined as SMEs with a high-growth ambition in phase (2) and (3), leaving phases (1), (4) and (5) for later. Historical growth data from Graydon allowed us to identify potential target SMEs, which were subsequently approached by UNIZO representatives.

In parallel with the entrepreneur promotion and recruitment phase, we recruited and trained coaches to become familiar with our coaching program, particularly the associated theoretical concepts and empirical tools (Step #2). For the recruitment of coaches, we made use of UNIZO's network of experienced coaches, who are either on the payroll of UNIZO or insourced as independent freelancers. We developed a training package consisting of four modules that coaches need(ed) to follow before they can advise entrepreneurs according to our methodology: (1) watch the online, pre-recorded training lecture in which the academic team explains the full methodology; (2) study the manuals, referred to as the *Enterprise Manual* and *Entrepreneur Manual*, explaining the different empirical tools; (3) join another coach already familiar with the methodology during a coaching track; and (4) regularly participate in the bi-monthly learning network gatherings to share experiences, discuss cases, and receive project updates and insights from the partners (i.e., the academic team and UNIZO).

In the first two years, we recruited over 500 entrepreneurs, and we are moving toward 900 at the end of year four. No less than 43% of the enterprises have fewer than 5 employees, and 74% have fewer than 10. As many as 84% of participants are owners or co-owners of their venture, and 31% are female. Approximately 14% of the enterprises indicate an acute need for help, but all decided to participate because they were looking for advice and support.

6. The Coaching Service

The core of our methodology is the evidence-based coaching service, which consists of four conversations of about three hours each between the coach and entrepreneur. On top of that, the coach needs, on average, two to three hours to prepare for each conversation. What makes our project very different from what commercial consultancies offer in the marketplace is the research-based evidence collected along the way that is the key input for these conversations (see below). This implies that that we strike two flies with one stone. On the one hand, scientific rigor is used to collect information that feeds into personal and strategic advice, and hence practical relevance. On the other hand, our collaboration with practice implies that this relevance facilitates scientific rigor, as we collect rich data for academic research. The whole coaching service takes at least one month, but often three or four, and should ideally be completed within two to a maximum of four months. The outcome is an advisory report written by the coach, identifying at least five to ten strategic and personal suggestions for the entrepreneur based on the quantitative results from the empirical tools, as well as the qualitative insights from the first three conversations. Below, we explain the purpose of each conversation, the tools involved, and the associated outcomes.

6.1. First Introduction: Intake

The purpose of the first conversation (Step #3) is for the coach to become better acquainted with the entrepreneur and her or his firm, and vice versa. Before they meet, the coach prepares her or himself by consulting the firm's website. An important element in our project is the involvement of Graydon. During the project, Graydon provides important demographic, financial and other available information about all participating SMEs, as well as many thousands of SMEs in the wider Flemish community (essential for the matched-pair analyses; see below). At this stage of the trajectory, by way of preparation of the first conversation, demographic, financial and other available information about each SME is provided by Graydon through an automatically-generated so-called *Trade Report* (*handelsrapport*, in Dutch), which the coach (and the academic team) can download from the Graydon website. This information includes, e.g., historical data regarding the SMEs' performance, as well as their sector and size.

The first intake conversation follows a semi-structured format. Several basic issues are discussed, such as the firm's core activities (e.g., products, services, and primary markets), organization (e.g., age, size, location(s), and structure) and transition phase (e.g., growth or decline). Depending on the transition phase, the coach further asks specific questions—for example, whether the entrepreneur has recently experienced severe setbacks (e.g., financial losses), whether the entrepreneur considers the firm in urgent need of help, whether the firm has sufficient funding available, and whether the situation is affecting the entrepreneur's personal life, too. If the entrepreneur is looking to be acquired, the coach will ask why the entrepreneur wants to sell the firm (e.g., retirement), whether the firm has already found a new owner (e.g., a family member or employee), and what s/he thinks is important during this transition period. The outcome of the first conversation is an *Intake Report*, for which a template is provided by the academic team. For each firm that participates in the project, the coach completes the *Intake Report*, and uploads the report through the online system of UNIZO so that they can monitor the coach's progress. The qualitative information is part of the project's data-collection effort. All text is included in the database, to be manually or automatically coded in due course.

6.2. Second Conversation: The Enterprise

The second conversation focuses entirely on the enterprise (Step #4). After their first conversation, the coach invites the entrepreneur to complete an online scan (referred to as the *Enterprise Scan*) developed by the academic team, with much feedback from practice in the co-creation team, with a series of questions about the firm. The *Enterprise Scan* takes about 30 to 45 min to complete. The questions are all drawn from prior academic research or are self-developed, and relate to a series of firm-level and environmental concepts. In most cases, entrepreneurs are asked to rate different statements belonging to a construct on a five or seven-point Likert scale (e.g., from 1 = 'strongly disagree' to 5 or 7 = 'strongly agree'). A complete overview of all concepts, scales and sources is provided in Table 1.

Table 1. Firm- and environmental-level constructs.

Construct.	Dimensions or Categories	Reference(s)
Organizational life cycle REMOVED *Coaches evaluated this assessment as too problematic, preferring to refer to qualitative narratives in the Intake Report.*	Stage 1: Existence Stage 2: Survival Stage 3: Success Stage 4: Renewal Stage 5: Decline	Lester et al. (2003)
Company environment	Market turbulence Competitive intensity Technological turbulence	Jaworski and Kohli (1993)
Macro-trend awareness	Importance attributed to a list of macro trends (e.g., Brexit, digitalization, and climate change) Development of a plan/strategy to respond to these macro trends	Self-developed
Decision-making/thinking logic	Effectuation Experimentation Pre-commitments Flexibility Affordable loss Causation	Sarasvathy (2001) Chandler et al. (2011)
Value strategy/strategic positioning	Product leadership Operational excellence Customer intimacy	Treacy and Wiersema (1993) Reimann et al. (2010)
Growth strategy	Ansoff's growth strategy Market penetration Market development Product development Diversification	Ansoff (1957)
Entrepreneurial orientation	Risk-taking Innovativeness Proactiveness Competitive aggressiveness Autonomy	Miller (1983) Covin and Slevin (1989) Hughes and Morgan (2007)
Exploration/exploitation	Exploration Exploitation	March (1991) Fernhaber and Patel (2012)
Stakeholder involvement	To what extent does your organization collaborate with the following stakeholders to innovate (a list of stakeholders is provided)	Zeng et al. (2010)

Table 1. *Cont.*

Construct.	Dimensions or Categories	Reference(s)
Reputation	The organization is a prominent player within its market segment The organization has high credibility Comparison of the organization's reputation, product/service offering, and reputation against the organization's most important competitor	Saxton, Todd (1997) Self-developed
Team	Productive conflict resolution Mature communication Goal clarity Common purpose Psychological safety Integration: Cross-functional interfaces Connectedness	Jaworski and Kohli (1993) Tekleab et al. (2009) Jansen et al. (2009) Final scale self-developed
International performance	Foreign sales as a Percentage of Total Sales (FSTS) Research and Development Intensity (RDI) Foreign Assets as a Percentage of Total Assets (FATA) Overseas Subsidiaries as a Percentage of Total Subsidiaries (OSTS) Top Managers' International Experience (TMIE) Psychic Dispersion of International Operations (PDIO)	Sullivan (1994)
Performance satisfaction	Performance satisfaction with Turnover Profit Average number of employees Foreign sales	Self-developed
Performance aspirations CHANGED *The horizon for the aspirations was changed from five to three years, as entrepreneurs considered a five-year look-ahead too difficult.*	Aspirations regarding Turnover Profit Average number of employees Foreign sales	Delmar and Wiklund (2008)
Performance expectations CHANGED *The horizon for the expectations was changed from five to three years, as entrepreneurs considered a five-year look-ahead too difficult.*	Expectations regarding Turnover Profit Average number of employees Foreign sales	Cassar (2006)

Note: REMOVED = a measure that was removed, and CHANGED = a measure that was changed. In *italics*, whenever appropriate, we briefly explain the reason for the adjustment.

For further information, we list the key reference regarding the original conceptual construct paper, as well as a publication providing the scale at hand for cases where the conceptual construct paper does not provide a measurement scale. We included two self-developed scales regarding macro-trend awareness and performance satisfaction, as we could not find reliable and valid measures in the extant literature that fit well with our purposes. Macro-trend awareness is not so much included for theoretical reasons, but rather for its relevance to practice. Performance satisfaction is interesting from an academic perspective as well, being an important outcome variable, as well as a subjective

proxy for performance. To construct a subjective assessment of enterprise performance, the survey asks the entrepreneur's evaluation of customers' overall satisfaction with her or his SME's products and/or services, the SME's reputation, and the SME's growth compared to competitors, as well as the entrepreneur's satisfaction with her or his SME's performance (i.e., in terms of revenues, profits, number of employees, and sales abroad) over the last three or five years, her or his growth wishes for the next year and next three or five years, plus her or his (more realistic) expectations regarding future performance of her or his SME.

After the entrepreneur has completed the *Enterprise Scan*, the data automatically becomes available to the academic team in the format of a four-page format-standardized *Enterprise Report*, providing the firm's scores on all measures. Per construct, the team calculates the overall score by averaging the scores across the items associated with that construct or, if the measure is multi-dimensional, each of that construct's dimensions (e.g., 1.2/5 for 'market turbulence', or 4.3/5 for 'operational excellence'). The *Enterprise Report* is then sent to the coach of that particular firm to prepare for their next conversation. We advise the coaches to continuously consult the manual of the *Enterprise Scan* to help her or him with the interpretation of the scores and possible attention-deserving linkages between the constructs. Regarding potential direct effects, the *Enterprise Report*'s results are quite straightforward, indicating either a 'good' or 'bad' score, or anything in-between—that is, a score that is likely to boost or hamper performance (particularly growth, as this project's main firm-level outcome variable), or is probably immaterial. For instance, the *Enterprise Scan* includes more than 40 items about different aspects of teamwork, such as whether the team has a clear and common purpose, can easily work together, and can quickly resolve conflicts. If the firm scores low on these aspects, this 'bad' set of scores points to an important issue for the coach to focus on during the final conversation.

Quite a few other measures are less straightforward, and can essentially only be interpreted in relation to the current performance of the firm or in interaction with scores for other constructs. For instance, enterprises are advised to focus on 'causation' when they are active in a stable environment, where they can easily predict and plan for the future, and 'effectuation' when the environment is more turbulent, implying that a trial-and-error approach is more appropriate (Vanderstraeten et al. 2020). In such cases, there are no 'good' or 'bad' scores for these (dimensions of) constructs in isolation. Whether they are 'good' or 'bad' depends upon their fit with other constructs—in our example, causation/effectuation vis-à-vis dynamic/stable environment. This is classic contingency logic. In the *Enterprise Report*, such instances of fit or misfit are explicitly identified if they really stand out as potentially outstanding or problematic. During the bi-monthly gatherings of the learning network, the interpretation of the (interaction between) scores and examples of cases are discussed among the coaches and the academic team. During the second conversation, the coach discusses the results of the *Enterprise Scan* with the entrepreneur, and keeps notes as input for the *Advisory Report*, which follows later (see below).

6.3. Third Conversation: The Entrepreneur

The third conversation (Step #5) moves closer to the heart of entrepreneurship, both literally and metaphorically speaking, by focusing on the entrepreneur as a person—as an individual with idiosyncratic attributes and attitudes. After the second conversation, the coach invites the entrepreneur to complete a second online scan (the *Entrepreneur Scan*), this time with a series of questions about the entrepreneur as a person. The *Entrepreneur Scan* takes about 45 min to an hour to complete. As with the *Enterprise Scan*, the questions are drawn from prior research or are self-developed, and all relate to the individual entrepreneur. An overview is provided in Table 2 (including the *BIATs*, as explained below).

Table 2. Entrepreneur-level constructs.

Construct.	Dimensions or Categories	Reference(s)
Personality: HEXACO Big Six	Honesty-Humility (H) Emotionality (E) Extraversion (X) Agreeableness (A) Conscientiousness (C) Openness to Experience (O)	Lee and Ashton (2004)
Personality: Explicit motives	Need for dominance Need for achievement Need for affiliation	McClelland (1965) Slabbinck et al. (2018)
Personality: BAS/BIS ADDED *Added as this is a fundamental pair of traits with substantive behavioral impact.*	Behavioral Approach System (BAS) Behavioral Inhibition System (BIS)	Carver and White (1994) Muehlfeld et al. (2013)
Personality: Dark Triad ADDED *Added as leadership research has revealed the Dark Triad's influence on behavior.*	Machiavellianism Narcissism Psychopathy	Paulhus and Williams (2002)
Intolerance	Intolerance of ambiguity Intolerance of uncertainty	Freeston et al. (1994) Carleton et al. (2007) McLain (2009)
Entrepreneurial self-efficacy	Searching Planning Marshalling Implementing (people) Implementing (financial)	Boyd and Vozikis (1994) McGee et al. (2009)
Leadership style CHANGED *A new leadership scale was added, because the coaches observed limited variance with the initial measure.*	Forceful leadership Enabling leadership New: Empowering leadership	Kaplan and Kaiser (2003) Arnold et al. (2000)
Emotional agility SHORTENED *The initial 30-item scale was reduced to 12 items, after a psychometric analysis, because the initial scale took too much space.*	Acceptation Committed action Defusion Mindfulness Self in context Values	Self-developed, inspired by David (2016)
Affect REMOVED *Coaches considered this concept to be of too little relevance.*	Positive affect Negative affect	Watson et al. (1988)
Social skills	Social perception Social adaptability Expressiveness Self-promotion Ingratiation	Baron and Markman (2000) Baron and Tang (2009)
Social support REMOVED *The measure using last name initial listing turned out to be too time-consuming to complete.*	Social support network Satisfaction with social support network	Cohen et al. (1985) Sarason et al. (1987)

Table 2. *Cont.*

Construct.	Dimensions or Categories	Reference(s)
Implicit motives	A BIAT to complement the explicit motives scale	Slabbinck et al. (2018)
Implicit HEXACO CHANGED *To further deepen academic knowledge, we made room for a Dark Triad BIAT.*	A BIAT to complement the explicit HEXACO scale New: A BIAT to complement the explicit Dark Triad scale	Self-developed
Implicit entrepreneurial self-efficacy	A BIAT to complement the explicit entrepreneurial self-efficacy scale	Self-developed

Note: REMOVED = a measure that was removed, SHORTENED = a scale that was substantially shortened, CHANGED = a measure that was changed, and ADDED = a novel measure that was added, all after the midway evaluation. In *italics*, whenever appropriate, we briefly explain the reason for the adjustment. BIAT = brief implicit association test.

The measures relate to action-theoretic frameworks in the entrepreneurship literature, such as those developed by Frese (2009); Frese and Gielnik (2014), and Newman et al. (2019), which link personality and motivational, affective and cognitive antecedents to action characteristics (including self-efficacy), and these in turn to entrepreneurial (success) outcomes, including growth and profitability. Again, for further information, we list the key reference regarding the original conceptual construct paper, as well as a publication providing the scale at hand for cases where the conceptual construct paper does not provide a measurement scale. To the *Entrepreneur Scan*, we added one self-developed scale, capturing emotional (or psychological) agility (or flexibility). This concept is new to entrepreneurship, originating from clinical psychology. Hence, we could not make use of an existing scale validated in the context of our target group.

Next to and on top of the *Entrepreneur Scan*, entrepreneurs are also asked to complete several Brief Implicit Association Tests (*BIATs*) to gain insight into their implicit (or unconscious) personality, which can only be revealed through indirect tests, of which *BIATs* are a prominent example (cf. Slabbinck et al. 2018). Take the example of implicit motives. People are driven by both explicit and implicit motives (e.g., the need for achievement, affiliation, and power). Implicit motives are shaped during early childhood, whereas explicit motives are formed only later, starting from puberty, when the person's motives are influenced by the environment (Schultheiss 2008). Key drivers of an individual's behavior are not only her or his explicit and implicit motives in isolation, but also the (in)congruence between both rankings. Before the midway redesign, the *BIATs* involved the HEXACO Big Six personality traits, entrepreneurial self-efficacy, and implicit motives. After that, we removed HEXACO, and added the Dark Triad. With the exception of the implicit motives BIAT, all *BIATs* are self-developed.

After the entrepreneur has completed both the *Entrepreneur Scan* and the *BIATs*, the data become available to the academic team, automatically for the *Entrepreneur Scan* and manually for the *BIATs*. Again, the academic team then creates another four-page *Entrepreneur Report* with the individual entrepreneur's personal scores (e.g., a 5.2/7 for explicit need for power, and 2.1/5 for emotionality), which is then sent to the coach of that particular entrepreneur to prepare for their next conversation. As with the *Enterprise Scan*, we advise the coaches to continuously consult the manual of the *Entrepreneur Scan* (which includes a discussion of the *BIATs*) to help her or him with the interpretation of the results. For instance, according to prior research, entrepreneurs with a dominant need for power tend to prefer more socially responsible and eco-sustainable strategies, but only if the entrepreneur does not consciously seek power (Hermans et al. 2017). During the bi-monthly gatherings, the interpretation of scores (in isolation and/or interaction) and examples of cases are discussed among the coaches and academic team.

During the third conversation, the coach discusses the scores of the *Entrepreneur Scan* and the *BIATs* with the entrepreneur, and keeps notes as an input for the *Advisory Report*, which is

presented and discussed during the fourth conversation (see next). Contrary to the conversation about the enterprise, coaches often invite entrepreneurs to have this conversation about her or his personality at another location than at the enterprise, because the issues on the table are much more personal and sensitive, and can quite easily turn emotional. At this point of the introduction of our example's design, we would like to make an extra remark, relating to issues of privacy, confidentiality, and desirability. Of course, during the intake conversation, all potential participants are explicitly informed about what they can expect, and that all information will be treated confidentially, with full protection of their privacy. Moreover, the entrepreneurs are urged to answer honestly, in their self-interest, in response to all questions. After all, meaningful advice cannot be expected if the answers have been dishonest. This issue, as well as that regarding privacy, is particularly relevant in the context of the *Entrepreneur Scan* and the *BIATs*, as these involve the entrepreneur as a person.

6.4. Fourth Conversation: A Tailor-Made Advice

Preparation of the fourth conversation (Step #6) involves bringing all pieces of information together as input for a final advice to the entrepreneur and her or his enterprise, seeking to offer the support that may increase the likelihood that s/he will succeed in reaching their ambition (whatever that may be). These pieces of information are: (1) data from Graydon; (2) qualitative intake narrative; (3) scores from the *Enterprise Scan*); (4) scores from the *Entrepreneur Scan*; (5) scores from the *BIATs*; and (6) any observation made along the way. Based on all of the scores (i.e., from the *Enterprise Scan*, the *Entrepreneur Scan*, and the *BIATs*), conversations and observations, the coach prepares an *Advisory Report* for the entrepreneur. This report consists of two main parts: a more objective description of the most striking findings (referred to as "mirrors"), and a list of several suggestions, both strategic as well as personal, for the entrepreneur and her or his firm, based on the whole coaching journey. During the fourth and final conversation, the coach discusses the *Advisory Report* with the entrepreneur, explaining the different findings and emphasizing a list of key suggestions. The entrepreneur is given a copy of the report for further reading and consultation.

A template of the *Advisory Report* was developed by the academic team, with extensive feedback from the learning network, resulting in about 25 so-called "mirror templates" that the coach can use in drafting their final reports. As explained above, the interpretation of the scores in isolation tends to be quite straightforward, whilst those that can only be interpreted in combination with other scores are more challenging. Contrary to the other three conversations, which first focus on distinct aspects of the enterprise and gradually move closer to the heart of the entrepreneur (from intake to the firm as an enterprise to the entrepreneur as a person), this final conversation offers the coach the opportunity to identify linkages between all previously-discussed topics. For instance, does the personality of the entrepreneur fit with the firm's business strategy? Does the entrepreneur still fulfil the role that best suits her or his personality, which is necessary to move the firm forward into the future? Take the following example: an entrepreneur who is open to new experiences (*Entrepreneur Scan*) may be better suited to explore new business ideas (*Enterprise Scan*), especially when the environment is turbulent (*Enterprise Scan*), whereas someone who is more introverted (*Entrepreneur Scan*) may be a better fit to exploit ideas (*Enterprise Scan*) to pursue an operational excellence strategy (*Enterprise Scan*), particularly in a stable environment (*Enterprise Scan*).

Of course, with dozens of measures of dozens of (dimensions of) concepts, the potential number of linkages—and hence 'fits' and 'misfits'—is incomprehensible, certainly from the perspective of practice. To cut this Gordian knot, in the learning network, we agreed that the coach would select the ten most striking 'mirrors' as the steppingstone for giving feedback to the entrepreneur and for identifying particular areas of improvement. The 'mirror' is a key notion that is a clear product of the close interaction between academia and practice. In a 'mirror', the academic team identifies a prominent 'misfit' in the sense of a 'bad' interaction between two scores, either within the *Enterprise Scan* or the *Entrepreneur Scan/BIATs*, or across both units of analysis.

In the so-called *Mirror Report* (*spiegelrapport*, in Dutch), after combining and analyzing across all sources of data, a list of ten such misfits is included, interpreted, and discussed with the aim to offer advice as to what the SME and entrepreneur can do to avoid or correct each of these misfits. For instance, the *Mirror Report* may argue that the entrepreneur's personality (say, a very high score on conscientiousness) does not align well with her or his enterprise's strategy (say, a very high score on exploration). The advice can then be to either change strategy (into exploitation) or to recruit another member in the enterprise's management team with a fitting personality trait (say, with a very high score on openness to experience) to implement the current strategy. It is up to the coach, in light of her or his 'soft' knowledge of the SME and the entrepreneur, to decide to include or exclude each 'mirror' and the associated advice in the *Advisory Report*.

In the final *Advisory Report*, three types of advice are included: strategic, personal and operational. First, the strategic advice relates mostly to the firm and the way the entrepreneur manages the business as well as the organization. For instance, the coach may suggest that, given the nature of the environment and strategy, to focus on gaining new clients rather than streamlining current operations, or to further improve the current products rather than developing new markets. The coach may also suggest that the entrepreneur should hire her or his first employee to perform certain administrative or operational tasks, so that the entrepreneur can concentrate on expanding the business. Second, the personal advice involves the general well-being of the entrepreneur. Although the coaching service does not intend to change the entrepreneur's personality (even if that would be possible), the coach may suggest that s/he takes a course in, for instance, time management or mindfulness, or that the entrepreneur delegates certain tasks to employees or switches responsibilities with other team members so that s/he can better use her or his core strengths. Third, the coach might decide to offer operational advice based on her or his observations regarding quick operational wins. For example, the coach may offer tips related to marketing policy, service innovation, internationalization or accounting practices that offer the entrepreneur a quick path to improvement, and refer to other experts in the field for further follow-up.

7. Satisfaction and Follow-up Survey

After finishing the coaching service, we measure the participating entrepreneurs' satisfaction with the coaching track, as well as the assessed impact of participation through two surveys: the *Satisfaction Survey* and the *Follow-Up Survey* (Step #7). The *Satisfaction Survey* is administered shortly after the end of the coaching service, and the *Follow-Up Survey* about six months after the advisory report was discussed. With this pair of measures, we seek to capture a subjective assessment of the immediate and mid-term impact of participation, in terms of both behavioral change (of both the enterprise and the entrepreneur) and performance (of different types, such as employee growth and financial profitability). Methodologically, doing so implies a within-subject design, as this pair of measures gives a subjective evaluation of (a) the extent to which the involved enterprise and entrepreneur have implemented changes and (b) the degree of performance improvement (or change, as a decline cannot be excluded) triggered by these changes, both according to the subjective assessment of the entrepreneur.

The *Satisfaction Survey* is administered to the entrepreneur shortly after the final conversation with the coach about the *Advisory Report*, and takes about 10 min to complete. The participating entrepreneurs are asked to evaluate the overall coaching service on a scale from 1 to 10 (from 1 = 'terrible' to 10 = 'excellent'), whether the coaching had met their expectations (from 1 = 'did not meet my expectations at all' to 10 = 'totally met my expectations'), to what extent they would recommend the service to friends or colleagues (from 1 = 'not at all' to 10 = 'definitely'), and to what extent they have already put the suggestions of the *Advisory Report* into practice (from 1 = 'I haven't done anything with it yet and I don't intend to do so' to 4 = 'I have applied everything that I could'). This feedback is practically relevant for UNIZO and VLAIO.

From an academic perspective, the follow-up measure is more useful as a subjective assessment measure. About six months after the final conversation with the coach, the entrepreneur receives a short *Follow-Up Survey*, which takes about five minutes to complete. This time, the entrepreneurs are asked to indicate whether the coach offered them mostly strategic, personal and/or operational advice, to what extent they are satisfied with the advice (from 1 = 'not satisfied at all' to 5 = 'totally satisfied'), and (again) to what extent they have put the advice into practice (from 1 = 'I haven't done anything with it yet and I don't intend to do so' to 4 = 'I have applied everything that I could'). Furthermore, we ask to what extent the coaching service has already influenced the firm's performance, such as growth in revenues, profits, number of employees, and sales abroad (from 1 = 'very negative' to 5 = 'very positive'), and to what extent they expect that implementing the advice will influence performance in the coming years. Finally, we ask to what extent the coaching service has influenced the entrepreneur's personal growth, and her or his happiness at work and at home, and with her or his work-life balance. This provides subjective assessment measures that can be taken as the outcome yardsticks in a within-subject design.

8. Follow-up and Matched-Pair Analyses

As a first step, running within-subject analyses with the subjective assessment measures can offer insights into the potential effectiveness of the coaching service. However, through the lens of causal inference, such a design is rather weak, being essentially (a) correlational with (b) subjective performance assessment data only. Hence, we add two types of data to provide the opportunity to run two types of complementary analyses (Step #8). The first is that we tap into Graydon's data to append longitudinal objective performance information. As time passes, we can follow up on all participating SMEs' performance in terms of (different types of) growth, profitability, and survival. In so doing, we construct a panel dataset that combines the rich qualitative and quantitative data collected during the coaching service with post hoc performance information. We refer to this dataset as the *Stand-Alone Panel*. Although still not a randomized control trial, of course, this panel design offers the opportunity to apply advanced causal inference econometrics (e.g., by adding time lags and instrumental variables).

Second, we are in the process of constructing a matched-pair dataset. After deciding about meaningful matching criteria (such as sector, size, and profitability at *t*, where *t* is the date at which the focal SME entered into the coaching service), we look in the Graydon database for an SME that is a matching twin in *t*, but that was not enrolled in the coaching program. Subsequently, we will keep track of all twins over time. We will do so by regularly updating each twin's performance by adding new information from Graydon. In this way, we construct a panel dataset, which we refer to as the *Twin Panel*, with parallel longitudinal information for twins, of which only one participated in the project. In a way, in doing so, we approximate a randomized controlled trial, albeit only imperfectly so. After all, although 50 per cent of the matched-pair dataset can be seen as the control group with SMEs that were not treated (that is, not having participated in the 'trial'), the selection of the treatment vis-à-vis control group has not been random. Notwithstanding this imperfection, these data will offer the opportunity to come even closer to causal inference through a between-subject design with SMEs that were 'treated' next to similar counterparts that were not.

9. Illustrative Analysis

The proof of the pudding is in the eating. Hence, to illustrate what can be done with data collected through a CARE design, we briefly present an example on the basis of data that have been gathered in the context of the *Stand-Alone Panel*, to date. The *Stand-Alone Panel* includes information from the *Follow-Up Survey* conducted about six months after the SMEs completed the coaching trajectory. By way of example, we take the item from this survey that asks them about their expectation of the extent to which their profit in the coming years will be influenced by their participation in the trajectory. With this item, we can run a tentative intervention impact analysis by estimating an OLS

regression model with the score on this item as the dependent variable. The item's score ranges from 1 ('Very negative influence') via 3 ('No influence') to 5 ('Very positive influence').

Regarding independent variables, we have a lengthy list to pick from, as is immediately clear from the above (see, e.g., Tables 2 and 3). For now, to keep our example manageable, we decided to focus on two key ones: one at the level of the SME from the **Enterprise Scan**, and one at the level of the individual from the **Entrepreneur Scan**. At the SME level, *Entrepreneurial Orientation* (*EO*) is one of the most established concepts in the entrepreneurship literature, as is *Entrepreneurial Self-Efficacy* (*ESE*) at the level of the individual entrepreneur. The foundational work of Miller (1983) and Covin and Slevin (1989) on *EO* is translated into a well-validated scale by Hughes and Morgan (2007). Similarly, we use McGee et al. (2009) scale of the *ESE* concept, as originally suggested by Boyd and Vozikis (1994). We refer to these pieces of work for further detail, including the underlying theoretical rationale. Here, we note that both scales are multidimensional. *EO* is composed of three dimensions (i.e., *Proactiveness*, *Innovativeness*, and *Risk-taking*), and *ESE* of six (i.e., regarding the five entrepreneurial activities of *Searching*, *Planning*, *Marshalling*, *People Managing*, *Finance Managing*, and *Venturing*). In our illustrative impact analysis, we explore whether and to what extent both aggregate *EO* and *ESE* measures, as well as the scores for the underlying elements, contribute to the effectiveness of the trajectory's coaching and advice at increasing SME long-term profitability.

Table 3. Example of an analysis: Expected impact on long-term profit in **Stand-Alone Panel**.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
ESE		−0.017		−0.015		
		(0.027)		(0.029)		
EO			−0.001	−0.0004		
			(0.001)	(0.001)		
Searching					0.024	
					(0.091)	
Planning					−0.346 ***	
					(0.103)	
Marshaling					0.132	
					(0.118)	
People managing					0.319 ***	
					(0.118)	
Finance managing					−0.079	
					(0.063)	
Venturing					0.031	
					(0.096)	
Proactiveness						−0.011
						(0.049)
Innovativeness						−0.009
						(0.049)
Risk-taking						0.012
						(0.051)
Age	−0.011	−0.011	−0.011	−0.011	−0.009	−0.010
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Gender	−0.124	−0.134	−0.130	−0.137	−0.142	−0.125
	(0.136)	(0.138)	(0.138)	(0.139)	(0.130)	(0.142)
Education	−0.138 **	−0.136 **	−0.141 **	−0.138 **	−0.065	−0.135 **
	(0.064)	(0.065)	(0.065)	(0.065)	(0.063)	(0.066)
Size	−0.048	−0.043	−0.048	−0.044	−0.020	−0.048
	(0.070)	(0.071)	(0.070)	(0.071)	(0.068)	(0.071)
Profit trend	0.0004	0.0004	0.0004	0.0004	0.001	0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Profit$_{t-1}$	−0.0003	−0.0003	−0.0003	−0.0003	−0.001	−0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0005)	(0.001)

Table 3. *Cont.*

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Venture age	−0.002	−0.001	−0.001	−0.001	−0.001	−0.001
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
Implementation	0.255 ***	0.255 ***	0.252 ***	0.253 ***	0.240 ***	0.255 ***
	(0.076)	(0.076)	(0.076)	(0.077)	(0.073)	(0.077)
Constant	4.126 ***	4.179 ***	4.143 ***	4.182 ***	3.315 ***	4.134 ***
	(0.551)	(0.559)	(0.555)	(0.562)	(0.725)	(0.575)
R2 (Adjusted R2)	0.37 (0.09)	0.37 (0.08)	0.37 (0.08)	0.37 (0.07)	0.48 (0.21)	0.37 (0.06)

Note: ESE = entrepreneurial self-efficacy, EO = entrepreneurial orientation, *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$; industry fixed effects are included; $n = 156$.

To filter out additional variance, we add a series of standard control variables. At the level of the individual entrepreneur, these are *Age*, *Gender*, and *Education*. For the SME level, we include *Profit* (both in year $t − 1$ and as a trend), *Sector, Size,* and *Venture Age*. We also show the relationship between application of advice and expected benefits from the advice with the variable *Implementation*, which is the item in the **Follow-Up Survey** asking for an assessment of the extent to which the advice is already implemented. Due to missing values and the time lag implied by the data collection delay associated with the **Follow-Up Survey**, our n is 156. We estimated six models: Model 1 includes the control variables only, Model 2 adds composite *ESE*, Model 3 adds composite *EO*, Model 4 adds both composite *ESE* and *EO*, Model 5 adds the six elements of *ESE*, and Model 6 adds the three elements of *EO*. The estimates of those models are presented in Table 3.

We provide all descriptive statistics and measurement detail in the Appendix A. For the sake of space, estimates for the dozens of *Sector* dummies are not reported (they are available upon request). From inspecting Table 3, we can conclude that four variables make a statistical difference (ignoring effect sizes, for now): *Planning*, *People Managing*, *Education,* and *Implementation*. Hence, we can conclude that the coaching and advice is viewed as having a positive impact on the SME's long-term profitability by less educated entrepreneurs with high people managing confidence and low planning confidence, and in ventures where the advice is already being applied. An example of a takeaway could be that promoting an SME's long-term profitability through coaching and advice should either be targeted at SMEs and their entrepreneurs already scoring high on people management confidence, or must go hand in hand with investment in boosting people management confidence. Of course, this impact analysis is just an example, as there is much more that can be and should be done. However, the example illustrates the type of data and impact associated with a CARE design.

10. Conclusions

The central argument in this paper is that we, as an academic community, can and should invest in co-developing and co-implementing creative research designs that combine rigor and relevance, working closely together with practice. That is, we argue that rigor and relevance should not be seen as an either/or tradeoff, but that we rather must creatively experiment with research designs that are characterized by both rigor and relevance. Here, we should move as close as we can to causal inference, as only then will our scholarly findings be rigorous enough to convincingly inform practice. In so doing, in close collaboration with practice, we will be able to conduct research with a dual impact, on both academic knowledge accumulation and practical organizational innovation, in line with the plea of the grassroots Responsible Research for Business and Management RRBM movement (see McKiernan 2016).

To do so, we first introduced six elements which we referred to as Co-creative Action Research Experiments (CARE), which we argued are primordial to developing studies able to bridge the rigor-relevance bridge. Building on these six elements, we developed a research design (see Figure 1), which, so we argue, can be applied to any project willing to bridge the rigor-relevance gap. While discussing this research design, we applied CARE to the Entrepreneurship field through

the Ambition in Entrepreneurship (*AiE*) coaching service project, in which we collected, and are still collecting, an impressive stock and flow of data on Flemish SMEs and their entrepreneurs over time through a variety of sources: i.e., the **Intake Report**, the **Enterprise Scan**, the **Entrepreneur Scan**, the **BIATs**, the **Advisory Report**, the **Satisfaction Survey**, **Follow-Up Survey**, **Stand-Alone Panel** and **Twin Panel**. In the near future, we will analyze the data for three different purposes, at least: to monitor and improve the coaching service, to conduct academic research, and to inform policy-making on entrepreneurship and SMEs. In all, in so doing, we will contribute to both practice (SMEs and society) and science, with a rigorous causal inference research design that is solution-oriented, being co-created and co-executed with key representatives from practice. Below, we briefly offer a few examples of what we did with the data through the lens of this set of three purposes.

Firstly, the **Satisfaction Survey** and **Follow-Up Survey** provide the information needed to immediately (just after ending the program participation) and quickly (after about six months) monitor whether entrepreneurs are satisfied with the coaching service, and to what extent they have applied the coach's advice into practice. This way, we can identify general areas of improvement, as well more specific issues that deserve further attention. For instance, generally, during the bi-monthly gatherings with the coaches at the early stages of the project, we discussed what concrete and hands-on suggestions could be offered as 'low-hanging fruit' or quick wins to support the participating entrepreneurs 'on the spot'. Specifically, for example, when we notice that an entrepreneur is not satisfied about a particular coach, we started to refer this coach to highly experienced and well-evaluated colleagues who may provide mentorship. Occasionally, the profile of the coach and that of entrepreneur simply failed to match well. Additionally, therefore, we suggested that UNIZO pays extra attention to their matching process.

Secondly, the rich variety of data offers ample opportunities to contribute to a wide range of academic fields, such as entrepreneurship, (behavioral) strategy, (personality) psychology, leadership, finance, marketing, and quite a few more. The key is that we can combine different types of data—for instance, coded information from the intake reports, survey measures from the scans, implicit scores from the BIATs, and financial information from Graydon. Apart from their multidisciplinary and theoretical richness, our data have a number of methodological strengths worth emphasizing. First, their multi-source nature implies that important biases are either missing, or can be carefully examined and/or corrected for. Examples include common-method variance and social desirability bias (e.g., Chang et al. 2010). Second, due to the longitudinal build-up, tracing SMEs over time after their participation, in combination with the matched-pair design, we can engage in rigorous causal inference. For instance, we can carefully examine the motivation-aspiration-performance chain (e.g., Hessels et al. 2008). Third, the action research focus on intervention comes close to a field experiment, further enhancing the causal inference strength. By coding the coach's advice in relation to the SMEs' extent of follow-up, we can explore how different types of interventions have different behavioral and performance consequences under different circumstances.

Thirdly, jointly, all these insights serve as input for formulating advice regarding policy-making aimed at stimulating different aspects of entrepreneurship. Here, we can run tailor-made analyses to look for answers to specific policy-related questions. For instance, what type of intervention is particularly effective to turn SMEs into gazelles, and what regulation-related bottlenecks do SMEs have to deal with on their journey toward high and sustained growth? In so doing, our design combines rigor and relevance, the latter not only from the perspective of the individual entrepreneur, but also through the macro lens of the government's policy-making.

As is the case for any research design, ours also has weaknesses. We would like to emphasize two, which are related: selection bias and information imbalance. First, selection bias is inherent to CARE. A matched-pair design can partly overcome this weakness, but creating twins is a complex process where perfect one-on-one matching is most likely never feasible. Second, applying the CARE principles resulted in four SME groups: (1) contacted and participated in program (i.e., the treatment group), (2) contacted but not participating in program, (3) not contacted and not included as a twin

(and thus not participated in program), and (4) not contacted but included as a twin. This caused an imbalance in information: We know much more about the SMEs that participated in the program compared to those that did not participate, and we only have limited information about the matched twins. We should thus be very cautious with conclusions that are based on data that we do not have for all companies.

Despite these weaknesses, our CARE-design can serve as a starting point for future research. For example, a successful CARE design requires a multidisciplinary approach to understand the ever-increasing complexity of our surrounding world (Nopens et al. 2019). Thus, rather than a myopic business focus, our CARE design integrates insights from a wide variety of disciplines in, e.g., economics, psychology and Sociology. Though commendable, future CARE designs should go further and move to deepen interdisciplinarity. The interdisciplinarity in our CARE design is limited to a few disciplines in the social sciences. Yet, next to social sciences, future CARE designs may also integrate insights, methods, and models from the arts and humanities, and natural sciences. Successful solutions to solve most of today's business and (wicked) societal problems require an intensive integration from an extensive set of research disciplines (Persson et al. 2018). To illustrate, as our society urgently needs to reduce its ecological impact, more effort is needed to motivate key decision-makers to take decisions that move our companies in that direction (Etzion 2018). Social scientists are well positioned to recommend 'how' these decision-makers may be motivated to take sustainable (business) decisions. However, social scientists often lack knowledge on 'what' should best be done (Persson et al. 2018). For example, in order to reduce the ecological impact of farming and consumption, should a farmer be motived to make a shift in production method (e.g., from traditional to organic) or in type of agricultural produce (from cattle to beans and pulses, as way to support the needed shift from animal to vegetable protein)? Answers to these questions obviously come from scientists in the natural sciences in general and agricultural scientists in particular. This simplified example illustrates that future CARE designs can greatly benefit from an intensive and close collaboration between scientists and experts from various disciplines.

Furthermore, our initial CARE design was developed in close cooperation between research institutions (AMS and University of Antwerp), a sector (UNIZO), and a supporting government institution (VLAIO). This co-creation resulted undoubtedly in a design that is both (scientifically) rigorous and (practically) relevant. However, our research is in essence top-down oriented. Research participants, entrepreneurs and their SMEs, in our case, were only limitedly involved in the initial development of the research. This approach was somewhat corrected by the bi-monthly learning network meetings that guaranteed that feedback of research participants was gradually integrated in the research design. Yet future CARE designs could go further by already involving research participants during the conceptual and developmental stages of the research design to warrant that their expectations, insights and concerns are already reflected in the first draft of a research design. Higher involvement of research participants may also lead to higher engagement between scientists and practitioners, which may further boost the rigor and relevance of research (Phillips et al. 2019). Put more generally, future CARE projects may benefit from principles and practices that have already been adopted in citizen science projects, as this type of research has a long history of involving research participants in the design, execution and interpretation phases of a scientific research project (Phillips et al. 2019).

In addition, the actors in the design process were almost exclusively highly educated people with academics, SME owners, business consultants and coaches as principal actors. Blue collar workers and other people at the bottom of the pyramid were not, or only to a lesser extent, consulted. As these people often have practical and creative solutions to solve relevant and day-to-day (business) problems, their insights may also give rise to better research designs (Gupta 2020). Interestingly, research on grassroots innovation already provides a compelling framework to incorporate the needs of people at lower levels in companies, communities or societies (Gupta 2020). Thus, principles that

have already been applied in grassroots innovation research may serve as a good example to further improve the foundations of future CARE designs.

Finally, to conclude, we would like to briefly reflect upon the issue of epistemology (cf. Hlady-Rispal and Jouison-Laffitte 2014). In management, traditionally, deduction is associated with quantitative and induction with qualitative work, with abduction being a rare betwixt-and-between epistemology. Potentially, the research design template we suggest here does combine all three ideal-typical epistemologies. For instance, deduction can inform the identification of independent variables and intervention strategies, the qualitative information can be the fuel of rich induction, and post hoc interaction analyses can involve an element of abduction. As our *AiE* example revealed, aiming at full deduction is not to be recommended (the full range of the data is far too broad and varied for that), induction alone is incompatible with causal inference through intervention, and abduction implies missed opportunities to contribute systematically to scientific knowledge accumulation (e.g., in the form of pre-registered replication). All in all, we hope that with our suggested six CARE principles, the research design and its application to the *AiE* coaching service, we were able to show its promise to contribute to bridging the rigor-relevance gap.

Appendix A

Table A1. Descriptive statistics and measurement detail illustrative impact analysis.

Concept	Measurement Detail	Descriptive Statistics
ESE	=(searching * planning * marshalling * people managing * finance managing * venturing)/1000	Min = 0.2, Mean = 3.2, Max = 12.5, SD = 2.2
EO	=proactiveness * innovativeness * risk-taking	Min = 1.8, Mean = 53, Max = 327, SD = 48
Searching	How much confidence do you have in your ability to identify the need for my product or service? (1 = very low … 5 = very high)	Min = 1.67, Mean = 3.9, Max = 5, SD = 0.70
Planning	How much confidence do you have in your ability to estimate customer demand? (1 = very low … 5 = very high)	Min = 1.75, Mean = 3.4, Max = 5, SD = 0.66
Marshaling	How much confidence do you have in your ability to run a network? (1=very low … 5 = very high)	Min = 2, Mean = 3.6, Max = 5, SD = 0.62
People managing	How much confidence do you have in your ability to train employees? (1=very low … 5 = very high)	Min = 1.5, Mean = 3.6, Max = 5, SD = 0.55
Finance managing	How much confidence do you have in your ability to read and interpret financial statements? (1 = very low … 5 = very high)	Min = 1, Mean = 3.5, Max = 5, SD = 0.87
Venturing	In general, starting a business is (1 = worthless … 5 = worthwhile)	Min = 2, Mean = 4.6, Max = 5, SD = 0.60
Proactiveness	The venture typically (1 = responds to actions competitors initiated … 7 = initiates actions to which competitors respond)	Min = 1.3, Mean = 4.3, Max = 7, SD = 1.36
Innovativeness	In the last three years, changes to product(lines) or services have been (1 = limited … 7 = drastic)	Min = 1, Mean = 3.4, Max = 7, SD = 1.41

Table A1. *Cont.*

Concept	Measurement Detail	Descriptive Statistics
Risk-taking	In general, the venture tends to favor (1 = low-risk projects, with normal profits . . . 7=high-risk projects, with a chance of high profits)	Min = 1, Mean = 3, Max = 7, SD = 1.32
Age	What is your age?	Min = 24, Mean = 44, Max = 70, SD = 8.3
Gender	0 = Male, 1 = Female	32% female
Education	1 = no school, 2 = primary school, 3 = secondary school, 4 = bachelor's degree, 5 = master's degree, 6 = Ph.D.	Min = 2, Mean = 4, Max = 6, SD = 0.86
Size	Number of employees (1 = 1–4; 2 = 5–9; 3 = 10–19; 4 = 20–49; 5 = 50–99; 6 = 100–199)	Min = 1, Mean = 1.8, Max = 6, SD = 1.04
Profit trend	Coefficient on year in, for each venture separately, a linear model on profit over the last four years.	Min = −743, Mean = 0.15, Max = 624, SD = 107
Profit$_{t-1}$	Annual profit in the most recent year/1000	Min = −164, Mean = 75, Max = 3800, SD = 280
Venture age	Year of founding minus 1970	Min = −43, Mean = 32, Max = 47, SD = 12.5
Industry	Two-digit NACE code	40 different industries
Implementation	To what extent have you already implemented the advice? (1 = I have not done anything yet, and I am not planning to, 2 = I have not done anything yet, but I am planning to, 3 = I have implemented a couple things, but not yet all, 4 = I have already applied all I can in my venture)	Min = 1, Mean = 2.9, Max = 4, SD = 0.67
Expected impact on long-term profit	To what extent do you think that the consulting project will still have an influence on profit in the coming years? (1=very negative, 3 = no influence, 5 = very positive)	Min = 1, Mean = 3.7, Max = 5,SD = 0.59

Note: ESE = entrepreneurial self-efficacy, EO = entrepreneurial orientation, * as the symbol for multiplication.

Table A2. Correlations illustrative impact analysis.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. ESE																			
2. EO	0.23																		
3. Searching	0.60	0.33																	
4. Planning	0.71	0.28	0.43																
5. Marshaling	0.65	0.26	0.46	0.53															
6. People managing	0.59	0.15	0.29	0.30	0.36														
7. Finance managing	0.54	−0.06	0.00	0.30	0.12	0.23													
8. Venturing	0.42	0.11	0.28	0.26	0.35	0.33	0.11												
9. Proactiveness	0.28	0.66	0.31	0.36	0.31	0.08	0.03	0.13											
10. Innovativeness	0.19	0.71	0.34	0.26	0.19	0.05	−0.05	0.02	0.41										
11. Risk-taking	0.23	0.80	0.26	0.25	0.21	0.19	−0.03	0.15	0.40	0.42									
12. Age	0.02	−0.03	0.03	0.10	−0.05	−0.03	0.07	0.07	0.08	0.03	−0.09								
13. Gender	−0.16	−0.10	−0.11	−0.17	−0.18	−0.10	−0.01	−0.13	−0.07	−0.11	−0.18	0.05							
14. Education	−0.05	−0.09	0.03	0.06	−0.05	−0.13	0.01	−0.09	−0.03	0.01	−0.07	0.00	0.16						
15. Size	0.09	0.05	−0.10	0.09	0.07	0.20	0.09	−0.00	−0.02	−0.01	0.04	0.03	−0.21	0.08					
16. Profit trend	0.07	0.14	0.11	0.05	−0.09	0.02	0.10	−0.03	0.05	0.20	0.14	0.09	0.01	0.09	−0.09				
17. Profit$_{t-1}$	−0.03	0.13	0.10	0.07	−0.01	−0.05	−0.09	−0.06	0.10	0.04	0.13	−0.01	−0.12	0.09	0.08	0.04			
18. Venture age	0.10	0.09	0.15	0.03	0.18	0.06	−0.04	0.11	0.20	−0.07	0.08	−0.11	0.01	−0.09	−0.34	−0.01	−0.12		
19. Implementation	0.08	0.06	0.08	0.12	0.17	−0.06	0.05	−0.12	0.10	0.15	−0.04	0.07	−0.01	0.02	−0.05	0.04	−0.01	−0.10	
20. Expected impact on long-term profit	0.07	−0.03	0.04	−0.07	0.13	0.10	0.01	0.07	−0.01	−0.01	−0.05	−0.06	−0.04	−0.16	−0.07	−0.04	−0.01	0.08	0.31

Note: ESE = entrepreneurial self-efficacy, EO = entrepreneurial orientation.

References

Aguinis, Herman, and Jeffrey R. Edwards. 2014. Methodological Wishes for the Next Decade and How to Make Wishes Come True. *Journal of Management Studies* 51: 143–74. [CrossRef]

Ansoff, Igor H. 1957. Strategies for diversification. *Harvard Business Review* 35: 113–24.

Antonakis, John, Samuel Bendahan, Philippe Jacquart, and Rafael Lalive. 2010. On Making Causal Claims: A Review and Recommendations. *The Leadership Quarterly* 21: 1086–120. [CrossRef]

Arnold, Josh A., Sharon Arad, Jonathan A. Rhoades, and Fritz Drasgow. 2000. The empowering leadership questionnaire: The construction and validation of a new scale for measuring leader behaviors. *Journal of Organizational Behavior* 21: 249–69. [CrossRef]

Baron, Robert A., and Gideon D. Markman. 2000. Beyond social capital: How social skills can enhance entrepreneurs' success. *Academy of Management Perspectives* 14: 106–16. [CrossRef]

Baron, Robert A., and Jintong Tang. 2009. Entrepreneurs' social skills and new venture performance: Mediating mechanisms and cultural generality. *Journal of Management* 35: 282–306. [CrossRef]

Bergh, Donald D., Barton M. Sharp, Herman Aguinis, and Ming Li. 2017. Is There a Credibility Crisis in Strategic Management Research? Evidence on the Reproducibility of Study Findings. *Strategic Organization* 15: 423–36. [CrossRef]

Birkinshaw, Julian, Mark P. Healey, Roy Suddaby, and Klaus Weber. 2014. Debating the Future of Management Research. *Journal of Management Studies* 51: 38–55. [CrossRef]

Boyd, Nancy, and George Vozikis. 1994. The influence of self–efficacy on the development of entrepreneurial intentions and actions. *Entrepreneurship Theory and Practice* 18: 63–77. [CrossRef]

Campbell, Benjamin A. 2013. Earnings effects of entrepreneurial experience: Evidence from the semiconductor industry. *Management Science* 59: 286–304. [CrossRef]

Carleton, Nicholas R., Peter J. Norton, and Gordon J. G. Asmundson. 2007. Fearing the unknown: A short version of the Intolerance of Uncertainty Scale. *Journal of Anxiety Disorders* 21: 105–17. [CrossRef] [PubMed]

Carver, Charles S., and Teri L. White. 1994. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology* 67: 319. [CrossRef]

Cassar, Gavin. 2006. Entrepreneur opportunity costs and intended venture growth. *Journal of Business Venturing* 21: 610–32. [CrossRef]

Chandler, Gaylen N., Dawn R. DeTienne, Alexander, McKelvie, and Troy V. Mumford. 2011. Causation and effectuation processes: A validation study. *Journal of Business Venturing* 26: 375–90. [CrossRef]

Chang, Sea-Jin, Arjen van Witteloostuijn, and Lorraine Eden. 2010. From the editors: Common method variance in international business research. *Journal of International Business Studies* 41: 178–84. [CrossRef]

Coad, Alex. 2009. *The Growth of Firms: A Survey of Theories and Empirical Evidence*. Cheltenham: Edward Elgar Publishing.

Cohen, Sheldon, Robin Mermelstein, Tom Kamarck, and Harry M. Hoberman. 1985. Measuring the Functional Components of Social Support. In *Social Support: Theory, Research and Applications*. Dordrecht: Springer, pp. 73–94. [CrossRef]

Covin, Jeffrey G., and Dennis P. Slevin. 1989. Strategic management of small firms in hostile and benign environments. *Strategic Management Journal* 10: 75–87. [CrossRef]

David, Susan. 2016. *Emotional Agility: Get Unstuck, Embrace Change, and Thrive in Work and Life*. New York: Penguin.

Dawson, Patrick. 2019. *Reshaping Change: A Processual Perspective*. New York: Routledge.

Delmar, Frederic, and Johan Wiklund. 2008. The effect of small business managers' growth motivation on firm growth: A longitudinal study. *Entrepreneurship: Theory & Practice* 32: 437–57.

DeTienne, Dawn R., and Gaylen N. Chandler. 2004. Opportunity Identification and Its Role in the Entrepreneurial Classroom: A Pedagogical Approach and Empirical Test. *Academy of Management Learning & Education* 3: 242–57. [CrossRef]

Etzion, Dror. 2018. Management for sustainability. *Nature Sustainability* 1: 744–49. [CrossRef]

Fernhaber, Stephanie A., and Pankaj C. Patel. 2012. How do young firms manage product portfolio complexity? The role of absorptive capacity and ambidexterity. *Strategic Management Journal* 33: 1516–39. [CrossRef]

Fiss, Peer C. 2011. Building Better Causal Theories: A Fuzzy Set Approach to Typologies in Organization Research. *Academy of Management Journal* 54: 393–420. [CrossRef]

Freeston, Mark H., Josée Rhéaume, Hélène Letarte, Michel J. Dugas, and Robert Ladouceur. 1994. Why do people worry? *Personality and Individual Differences* 17: 791–802. [CrossRef]

Frese, Michael. 2009. Towards a psychology of entrepreneurship: An action theory perspective. *Foundations and Trends® in Entrepreneurship* 5: 437–96. [CrossRef]

Frese, Michael, and Michael M. Gielnik. 2014. The psychology of entrepreneurship. *Annual Review of Organizational Psychology and Organizational Behavior* 1: 413–38. [CrossRef]

Gow, Ian D., David F. Larcker, and Peter C. Reiss. 2016. Causal Inference in Accounting Research. *Journal of Accounting Research* 54: 477–523. [CrossRef]

Gray, Barbara, and Jill Purdy. 2018. *Collaborating for Our Future*. Oxford: Oxford University Press (OUP). [CrossRef]

Gulati, Ranjay, Samuel B. Bacharach, and Peter A. Bamberger. 2007. Tent Poles, Tribalism, and Boundary Spanning: The Rigor-Relevance Debate in Management Research. *Academy of Management Journal* 50: 775–82. [CrossRef]

Gupta, Shaphali. 2020. Understanding the feasibility and value of grassroots innovation. *Journal of the Academy of Marketing Science* 48: 941–65. [CrossRef]

Hermans, Julie, Johanna Vanderstraeten, Marcus Dejardin, Dendi Ramdani, and Arjen van Witteloostuijn. 2014. L'entrepreneur ambitieux: État les lieux et perspectives (Ambitious Entrepreneurship: The current state of the art). *Revue de L'entrepreneuriat* 12: 43–70. [CrossRef]

Hermans, Julie, Hendrik Slabbinck, Johanna Vanderstraeten, Jacqueline Brassey, Marcus Dejardin, Dendi Ramdani, and Arjen van Witteloostuijn. 2017. The power paradox: Implicit and explicit power motives, and the importance attached to prosocial organizational goals in SMEs. *Sustainability* 9: 2001. [CrossRef]

Hessels, Jolanda, Marco Van Gelderen, and A. Roy Thurik. 2008. Entrepreneurial aspirations, motivations, and their drivers. *Small Business Economics* 31: 323–39. [CrossRef]

Hlady-Rispal, Martine, and Estèle Jouison-Laffitte. 2014. Qualitative research methods and epistemological frameworks: A review of publication trends in entrepreneurship. *Journal of Small Business Management* 52: 594–614. [CrossRef]

Hughes, Mathew, and Robert E. Morgan. 2007. Deconstructing the relationship between entrepreneurial orientation and business performance at the embryonic stage of firm growth. *Industrial Marketing Management* 36: 651–61. [CrossRef]

Jansen, Justin J. P., Michiel P. Tempelaar, Frans A. Van Den Bosch, and Henk W. Volberda. 2009. Structural differentiation and ambidexterity: The mediating role of integration mechanisms. *Organization Science* 20: 797–811. [CrossRef]

Jaworski, Bernard J., and Ajay K. Kohli. 1993. Market orientation: Antecedents and consequences. *Journal of Marketing* 57: 53–70. [CrossRef]

Kaplan, Robert E., and Robert B. Kaiser. 2003. Developing versatile leadership. *MIT Sloan Management Review* 44: 19–26.

Khurana, Rakesh. 2010. *From Higher Aims to Hired Hands. From Higher Aims to Hired Hands*. Princeton: Princeton University Press, Walter de Gruyter GmbH. [CrossRef]

Kieser, Alfred, and Lars Leiner. 2009. Why the Rigour-Relevance Gap in Management Research Is Unbridgeable. *Journal of Management Studies* 46: 516–33. [CrossRef]

Lee, Kibeom, and Michael C. Ashton. 2004. Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research* 39: 329–58. [CrossRef]

Leitch, Claire. 2007. An Action Research Approach to Entrepreneurship. In *Handbook of Qualitative Research Methods in Entrepreneurship*. Cheltenham: Edward Elgar Publishing, pp. 144–69. [CrossRef]

Lester, Donald L., John A. Parnell, and Shawn Carraher. 2003. Organizational life cycle: A five-stage empirical scale. *The International Journal of Organizational Analysis* 11: 339–54. [CrossRef]

March, James G. 1991. Exploration and exploitation in organizational learning. *Organization Science* 2: 71–87. [CrossRef]

Maula, Markku, and Wouter Stam. 2019. Enhancing Rigor in Quantitative Entrepreneurship Research. *Entrepreneurship Theory and Practice*, 1–32. [CrossRef]

McClelland, David C. 1965. Toward a theory of motive acquisition. *American Psychologist* 20: 321. [CrossRef]

McGee, Jeffrey E., Mark Peterson, Stephen L. Mueller, and Jennifer M. Sequeira. 2009. Entrepreneurial self—Efficacy: Refining the measure. *Entrepreneurship Theory and Practice* 33: 965–88. [CrossRef]

McKiernan, Peter. 2016. *A Vision for Responsible Research in Business and Management: Striving for Useful and Credible Knowledge*. Position paper. Preprint. Available online: https://strathprints.strath.ac.uk/59976/ (accessed on 28 September 2020).

McLain, David L. 2009. Evidence of the properties of an ambiguity tolerance measure: The Multiple Stimulus Types Ambiguity Tolerance Scale–II (MSTAT–II). *Psychological Reports* 105: 975–88. [CrossRef] [PubMed]

Meyer, Klaus E., Arjen Van Witteloostuijn, and Sjoerd Beugelsdijk. 2017. What's in a P? Reassessing Best Practices for Conducting and Reporting Hypothesis-Testing Research. *Journal of International Business Studies* 48: 535–51. [CrossRef]

Miller, Danny. 1983. The correlates of entrepreneurship in three types of firms. *Management Science* 29: 770–91. [CrossRef]

Muehlfeld, Katrin, Utz Weitzel, and Arjen van Witteloostuijn. 2013. Fight or freeze? Individual differences in investors' motivational systems and trading in experimental asset markets. *Journal of Economic Psychology* 34: 195–209. [CrossRef]

Newman, Alexander, Martin Obschonka, Susan Schwarz, Michael Cohen, and Ingrid Nielsen. 2019. Entrepreneurial self-efficacy: A systematic review of the literature on its theoretical foundations, measurement, antecedents, and outcomes, and an agenda for future research. *Journal of Vocational Behavior* 110: 403–19. [CrossRef]

Nopens, Ingmar, Kim Ragaert, Katrien Verleye, Hendrik Slabbinck, Iris Vermeir, Kasper Ampe, Erik Paredis, and Jan Arends. 2019. Interdisciplinary Research Urgently Needs Facilitation. Available online: https://www.linkedin.com/pulse/interdisciplinary-research-urgently-needs-ingmar-nopens (accessed on 24 September 2020).

Paulhus, Delroy L., and Kevin M. Williams. 2002. The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality* 36: 556–63. [CrossRef]

Persson, Johannes, Alf Hornborg, Lennart Olsson, and Henrik Thorén. 2018. Toward an alternative dialogue between the social and natural sciences. *Ecology and Society* 23: 14:1–14:11. [CrossRef]

Phillips, Tina B., Heidi L. Ballard, Bruce V. Lewenstein, and Rick Bonney. 2019. Engagement in science through citizen science: Moving beyond data collection. *Science Education* 103: 665–90. [CrossRef]

Prowse, Martin, and Laura Camfield. 2013. Improving the quality of development assistance: What role for qualitative methods in randomized experiments? *Progress in Development Studies* 13: 51–61. [CrossRef]

Reason, Peter. 2006. Choice and Quality in Action Research Practice. *Journal of Management Inquiry* 15: 187–203. [CrossRef]

Reeb, David M., Mariko Sakakibara, and Ishtiaq P Mahmood. 2012. From the Editors: Endogeneity in International Business Research. *Journal of International Business Studies* 43: 211–18. [CrossRef]

Reimann, Martin, Oliver Schilke, and Jacquelin S. Thomas. 2010. Toward an understanding of industry commoditization: Its nature and role in evolving marketing competition. *International Journal of Research in Marketing* 27: 188–97. [CrossRef]

Sarason, Irwin G., Barbara R. Sarason, Edward N. Shearin, and Gregory R. Pierce. 1987. A brief measure of social support: Practical and theoretical implications. *Journal of Social and Personal Relationships* 4: 497–510. [CrossRef]

Sarasvathy, Saras D. 2001. Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency. *Academy of Management Review* 26: 243–63. [CrossRef]

Saxton, Todd. 1997. The effects of partner and relationship characteristics on alliance outcomes. *Academy of Management Journal* 40: 443–61.

Schultheiss, Oliver C. 2008. Implicit motives. In *Handbook of Personality Psychology: Theory and Research*. Edited by Oliver P. John, Richard W. Robins and Pervin Lawrence A. New York: Guilford Press, pp. 603–33.

Slabbinck, Hendrik, Arjen van Witteloostuijn, Julie Hermans, Johanna Vanderstraeten, Marcus Dejardin, Jacqueline Brassey, and Dendi Ramdani. 2018. The added value of implicit motives for management research Development and first validation of a Brief Implicit Association Test (BIAT) for the measurement of implicit motives. *PLoS ONE* 13: e0198094. [CrossRef]

Starbuck, William H. 2016. 60th Anniversary Essay. *Administrative Science Quarterly* 61: 165–83. [CrossRef]

Sullivan, Daniel. 1994. Measuring the degree of internationalization of a form. *Journal of Business Studies* 25: 325–42. [CrossRef]

Tekleab, Amanuel G., Narda R. Quigley, and Paul E. Tesluk. 2009. A longitudinal study of team conflict, conflict management, cohesion, and team effectiveness. *Group & Organization Management* 34: 170–205.

Toh, Rex S., Eunkyu Lee, and Michael Y. Hu. 2006. Social desirability bias in diary panels is evident in panelists' behavioral frequency. *Psychological Reports* 99: 322–34. [CrossRef] [PubMed]

Treacy, Michael, and Fred Wiersema. 1993. Customer intimacy and other value disciplines. *Harvard Business Review* 71: 84–93.

Tsui, Anne S. 2013a. 2012 Presidential Address—On Compassion In Scholarship: Why Should We Care? *Academy of Management Review* 38: 167–80. [CrossRef]

Tsui, Anne S. 2013b. The Spirit of Science and Socially Responsible Scholarship. *Management and Organization Review* 9: 375–94. [CrossRef]

Van Burg, Elco, Joep Cornelissen, Wouter Stam, and Sarah Jack. 2020. Advancing Qualitative Entrepreneurship Research: Leveraging Methodological Plurality for Achieving Scholarly Impact. *Entrepreneurship Theory and Practice*, 1–18. [CrossRef]

Van de Ven, Andrew H., and Paul E. Johnson. 2006. Knowledge for Theory and Practice. *Academy of Management Review* 31: 802–21. [CrossRef]

van Witteloostuijn, Arjen. 2015. Toward Experimental International Business. *Cross Cultural Management: An International Journal* 22: 530–44. [CrossRef]

van Witteloostuijn, Arjen. 2016. What happened to Popperian falsification? Publishing neutral and negative findings. *Cross Cultural & Strategic Management* 23: 481–508.

Vanderstraeten, J., J. Hermans, Arjen van Witteloostuijn, and Marcus Dejardin. 2020. SME innovativeness in a dynamic environment: Is there any value in combining causation and effectuation? *Technology Analysis & Strategic Management*. [CrossRef]

Walker, Richard M., Yanto Chandra, Jiasheng Zhang, and Arjen Witteloostuijn. 2019. Topic Modeling the Research-Practice Gap in Public Administration. *Public Administration Review* 79: 931–37. [CrossRef]

Watson, David, Lee Anna Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* 54: 1063–70. [CrossRef]

Watts, Duncan J. 2017. Should Social Science Be More Solution-Oriented? *Nature Human Behaviour* 1: 15. [CrossRef]

Zeng, Saixing, Xuemei Xie, and Chiming Tam. 2010. Relationship between cooperation networks and innovation performance of SMEs. *Technovation* 30: 181–19. [CrossRef]