



Navigating the validity tradeoffs of entrepreneurship research experiments: A systematic review and best-practice suggestions

Denis A. Grégoire^{a,*}, Julia K. Binder^b, Andreas Rauch^c

^a HEC Montréal, Department of Entrepreneurship and Innovation, 3000, Chemin de la Côte-Sainte-Catherine, Montréal (Québec), Canada H3T 2A7

^b École Polytechnique Fédérale de Lausanne (EPFL), College of Management of Technology, Chair of Entrepreneurship and Technology Commercialization, Odyssea 3.17, Station 5, CH-1015 Lausanne, Switzerland

^c The University of Sydney Business School, Rm 4231, Abercrombie Building, University of Sydney, NSW 2006, Australia

ARTICLE INFO

Keywords:

Entrepreneurship
Research methods
Experiments

ABSTRACT

Experimental methods provide important advantages for advancing academic understanding of entrepreneurship. Yet, the complex and comingled relationships between some of entrepreneurship's key characteristics pose thorny methodological challenges to entrepreneurship researchers – notably to negotiate important tradeoffs between the ideals of external and construct validity. To facilitate the sound mobilization of experimental methods in entrepreneurship research, we present an overview of critical validity challenges plaguing entrepreneurship research experiments and assess the validation practices mobilized in 144 studies using such methods. Building on these findings, we develop a practical guide of actionable validation strategies to help experimenters navigate the above tradeoffs and conduct entrepreneurship research experiments that are realistic, theoretically meaningful, and that help establish the causal effects of their focal variables. By doing so, we contribute a set of pragmatic means to support the mobilization of experimental methods for advancing entrepreneurship research.

Executive summary: Experimental methods hold important advantages for advancing the theoretical and practical understanding of entrepreneurial action. The power of research experiments lies in their ability to yield convincing evidence supporting the causal effects of the factors they investigate.

Unfortunately, entrepreneurship research experiments are often criticized for offering pale copies of *both* the realities they attempt to model *and* the theoretical constructs they aim to study. The entrepreneurial process and its related activities count many intertwined characteristics – such as radical uncertainty, temporal dynamics, high personal stakes and other constraints – that can prove difficult to integrate in experimental studies. Moreover, these characteristics' comingled relationships raise important tensions with experimental methods' core principle of surgically focusing on the causal effects of a few manipulated variables. As a result, entrepreneurship research must overcome a number of validity challenges and tradeoffs in order to successfully leverage experimental methods and offer findings that convincingly support the causality of their theoretical developments.

Compounding these difficulties, the guidance typically offered in research methods textbooks tends to focus on generic issues that are altogether separate from the specific challenges of conducting valid entrepreneurship research experiments. Because of this, many research manuscripts mobilizing experimental methods arrive at the review process with important shortcomings. This threatens the field's knowledge accumulation and typically calls for authors to refine their work and collect additional data.

* Corresponding author.

E-mail addresses: denis.gregoire@hec.ca (D.A. Grégoire), julia.binder@epfl.ch (J.K. Binder), andreas.rauch@sydney.edu.au (A. Rauch).

<https://doi.org/10.1016/j.jbusvent.2018.10.002>

Received 1 May 2017; Received in revised form 30 September 2018; Accepted 10 October 2018

Available online 24 October 2018

0883-9026/ © 2018 Elsevier Inc. All rights reserved.

To help researchers face these challenges, we offer a three-part compendium focused on bolstering the validity of entrepreneurship research experiments. First, we complement the generic observations of research methods monographs by presenting an overview of the validity tradeoffs inherent to mobilizing experimental methods in entrepreneurship research – and of strategies for addressing these.

Second, and to demonstrate that overcoming these challenges is not a trivial task, we conduct a structured literature review of the external and construct validation strategies deployed in a comprehensive corpus of 144 relevant articles published in both entrepreneurship-focused journals and broader generalist journals from the applied behavioral and social sciences. Among the most positive elements emerging from our analyses, we note upward trends in the mobilization of pilot tests, in the conduct of multiple experiments within single papers as well as in the use of parallel studies using different data collection methods. We also observe increased efforts to explain the practical relevance of experimental findings. To our surprise, however, our analyses uncover downward trends in the mobilization of pre-tests to examine the research material's external validity and representativeness, as well as the continued publication of studies that do not report empirical evidence regarding their focal manipulations' construct validity. This is concerning. Such practices undermine these experiments' abilities to yield convincing evidence in support of their theorized causal effects – and for what these imply for fostering entrepreneurial action.

Third, and in light of the results obtained, we develop a practical guide of actionable strategies for navigating the validity tradeoffs of entrepreneurship research experiments. Grouping these strategies together allows us to systematize recommendations typically found across various methods monographs, thereby offering an overall scheme to help entrepreneurship researchers navigate the validity tradeoffs inherent to conducting realistic and theoretically meaningful experiments. To make our recommendations as actionable as possible, we develop an extensive step-by-step guide of design and assessment strategies relevant to entrepreneurship research experiments. The guide spans the entire research process – from conceiving an experimental study to collecting, analyzing and reporting the results. The guide also includes a number of relevant exemplars from many different studies we analyzed.

By discussing these entrepreneurship-specific validity tradeoffs and combining that with an overview of strategies for addressing them, our study offers an entrepreneurship-centered synthesis that equips entrepreneurship researchers with the necessary tools for conducting experimental research that advances understanding of the causal effects supporting entrepreneurial action and its related phenomena.

JEL keywords

C9 Design of experiments
 C91 Laboratory, individual behavior
 C92 Laboratory, group behavior
 C93 Field experiments
 C8 Data collection and estimation methodology

1. Introduction

Validity is at the heart of experimental research.¹ In scientific endeavors, validity refers to “the approximate truth of an inference (Shadish et al., 2002: p. 34).” It concerns the extent to which researchers, educators and practitioners can be confident of a study's conclusions. But establishing the validity of an entrepreneurship research experiment can pose important methodological challenges.

The entrepreneurial process and its related activities count many intertwined characteristics – such as radical uncertainty, temporal dynamics, high personal stakes and other constraints – that often prove difficult to integrate in experimental studies (see Hsu et al., 2017a: p. 387). Moreover, these characteristics' comingled relationships raise important tensions with experimental methods' core principle of surgically focusing on the effects of a few manipulated variables. To obtain convincing evidence in support of the causal dynamics fostering entrepreneurial action, researchers must therefore establish that their experimental manipulations, material and procedures convincingly align with *both* the real-life phenomenon they attempt to model *and* the theoretical effects they aim to study (see Davidsson 2006: pp. 314–318; Grégoire et al., 2010: pp. 119–120). Navigating tradeoffs of external and construct validity is thus at the heart of ongoing efforts to conduct sound experimental research in entrepreneurship (cf. Aguinis & Lawal 2012: pp. 494–496).

¹ From the Latin roots for “strength”, the term *validity* concerns the “manifest power (of) arguments, reasons, principles, processes,” or other evidence and proofs “to impress themselves on others as right and well-grounded (Webster's New Dictionary of Synonyms, 1984),” as “having legal efficacy or force,” as being “justifiable,” “logically correct” or “appropriate (Merriam-Webster's Online Dictionary, 2017).”

Compounding these difficulties, the guidance typically offered in research methods textbooks tends to focus on generic issues of research design, internal validity and/or statistical analysis (cf. Alferes, 2012; Brewer, 2000; Kirk, 2013; Maxwell and Delaney, 2004; McClelland, 2000; Mead et al., 2012; Shadish et al., 2002; Smith, 2000; West et al., 2000). Even recent reviews of entrepreneurship research experiments offer scant recommendations for addressing validity challenges. For instance, Kraus et al. (2016) observe that “(a) frequent methodological issue in experimental entrepreneurship research appears to be the quality and validity of manipulation in experimental settings (p. 971),” but their text does not provide further details on how to address this issue. For their part, Hsu et al. (2017a) draw attention to a number of relevant considerations, but “as a primer for those interested but not familiar with these methods (p. 379),” their text was never meant to offer exhaustive recommendations specific to the validation of entrepreneurship research experiments. Yet given the challenges’ importance and the dearth of entrepreneurship-specific advice, many research manuscripts mobilizing experiments arrive at the review process with important validity shortcomings. This threatens the field’s knowledge accumulation and typically calls for authors to refine their work and collect additional data.

To help researchers overcome these challenges, we offer a three-part compendium focused on bolstering the validity of entrepreneurship research experiments. First, we complement the generic observations of research methods monographs by presenting an overview of the validity challenges inherent to mobilizing experimental methods in entrepreneurship research. Second, and to demonstrate that overcoming these challenges is not a trivial task, we conduct a structured literature review of the external and construct validation strategies deployed in 144 relevant articles. Third, and in light of the results obtained, we develop a practical guide of actionable strategies for navigating the validity tradeoffs of entrepreneurship research experiments. By doing so, our work’s value-added contributions are 1) to draw attention to the particular nature of validity challenges in entrepreneurship research experiments, 2) to empirically document the more salient shortcomings in this regard, and 3) to offer guidance on relevant strategies for addressing these validity threats. Through it all, we hope to facilitate the sound mobilization of experimental methods to advance academic understanding of entrepreneurial phenomena.

2. The validity challenges of entrepreneurship experiments

By comparison to other endeavors typically studied in the applied human and social sciences, entrepreneurial action and its related activities generally involve particularly high levels of ambiguous signals, incomplete information and radical uncertainty, particularly high levels of personal stakes, commitments, investments, affective attachments, passions and other emotional dynamics, as well as particularly high levels of task demands, information overload, time pressures and other constraints (cf. Baron 1998: pp. 278–9; Davidsson, 2016; Gartner, 2001; Hsu et al., 2017a: p. 387; Katz and Gartner, 1988; McMullen and Dimov, 2013; Townsends et al., 2018). Needless to say, these characteristics are not unique to entrepreneurial endeavors: many human activities ranging from pursuing a promising career in the Arts to engaging in politics, managing a large corporation or conducting scientific research can exhibit the same characteristics. Yet their particular salience and co-occurrence in entrepreneurial pursuits pose important validity challenges to experimenters.²

2.1. Validity tradeoffs in entrepreneurship research experiments

The more salient characteristics of entrepreneurship are intrinsically co-mingled with one another. For instance, the *ex-ante* uncertainty of entrepreneurial projects goes hand-in-hand with issues of incomplete information, ambiguous signals and the entrepreneurial process’ temporal unfolding, thus calling for high-levels of personal commitments and early investments from a variety of stakeholders, which in turn yields high levels of affective attachments and emotional dynamics. As a result, the design and conduct of sound entrepreneurship research experiments call for researchers to negotiate important tradeoffs between the methodological ideals of construct and external validity.

A first illustrative tradeoff concerns the transient nature of new venturing ideas and the entrepreneurial process’ undetermined outcomes. On the one hand, methodological ideals of external validity encourage the use of real-life cases in designing a study’s material (cf. Grégoire et al., 2010; Wood et al., 2017). On the other hand, the uncertain nature of entrepreneurial endeavors implies that in order to conduct a validly realistic experiment, the ultimate success of these stimuli should also remain indeterminate.

In experimental studies on the emergence of entrepreneurial pursuits, for instance, it would seem incongruous to examine participants’ ideation or evaluation efforts with respect to real-life venturing ideas that are at such an advanced stage of development and ubiquity that these ideas have become near-certain exploited opportunities. Imagine conducting an experiment (in 2018 or later) using manipulated scenarios based on market applications of the same 3D-printing technologies that emerged in the 1990s (see Shane and Venkataraman, 2000). These technologies’ applications are so common nowadays (see <https://3dprintingindustry.com/>) that participants’ processing of the experiment’s material would likely have more to do with the ubiquitous nature of the technologies’ application than with the study’s theoretically-intended manipulations. The same would apply in an experimental study of the decision to invest in a “hypothetical” new venture whose description reads just like that of [amazon.com](https://www.amazon.com/)®, [expedia](https://www.expedia.com/)® or [facebook](https://www.facebook.com/)® (and where there is no longer any ambivalence about the ventures’ potential, at least under the conditions prevailing at the time of writing). In all these situations, the very meaning of the experiments’ manipulations would depart from the theoretical intent and be re-interpreted in light of participants’ knowledge of what happened in real-life.

It follows that for all the benefits that can accrue from basing an experiment’s material on real-life cases, entrepreneurship

² We express our profound gratitude to three anonymous reviewers for encouraging us to better develop this point.

experimenters face a tradeoff of ensuring that their research material and procedures still retain a defensible degree of open-endedness – in the sense that the information presented remains ambivalent and incomplete, and that its outcomes/chances of success also remain normatively uncertain. Otherwise, participants' intuitive understanding of the research stimuli's plausibility and ubiquity would introduce confounding factors in the research. This would not only violate the study's external validity (in light of the uncertainty characterizing entrepreneurial pursuits), it would invalidate the manipulation's theoretical intent (construct validity), thereby undermining the experiment's entire set of causal inferences.³

A second illustrative tradeoff concerns participants' involvement with an experiment's material and procedures. Scholars and practitioners have long pointed out that entrepreneurial action involves high levels of personal immediacy and attachment (cf. Baron 1998: pp. 278–9; Cardon et al., 2005; McMullen and Kier, 2016). Entrepreneurs often have a lot of resources and personal stakes committed in their projects, with real possibilities of losing it all. But with the possible exception of natural quasi-experiments and active deceptions of participants' own pursuits (which both pose other issues, see Hsu et al., 2017a), few ethically defensible strategies allow for conveying the same high levels of immediacy and attachment that characterize entrepreneurial efforts. In most cases, important ethical and practical concerns make it impossible to experimentally manipulate the real-life elements of entrepreneurs' projects.

For instance, it would not do to arbitrarily cut a venture's funding simply to prove that this could limit its growth (cf. Kerr et al., 2011), no more than it would be ethical to prevent a real-life entrepreneur from sleeping just to demonstrate the adverse effects that this would have on his/her venture's success (cf. Weinberger et al., 2018). Consequently, entrepreneurship experiments typically resort to hypothetical exercises, artificial scenarios, deceptions, role playing and/or stark contrasts in some manipulated factors that are all meant to approximate these otherwise important realities (see Hsu et al., 2017a: pp. 10–20). These strategies also have the advantage of allowing for studying phenomena that are otherwise infrequent – or whose onset is difficult to predict and/or investigate “live” and “in-situ.” In doing so, however, researchers implicitly assume that participants' potentially lessened engagement with their experiments' task(s), material and procedures will not meaningfully affect the obtained results – by comparison to what would ensue if the manipulations directly concerned participants' personal circumstances and/or venture(s). But for this assumption to hold, it becomes critical that participants remain attentive and engaged with the experiment's material and procedures. Otherwise, there is a risk that participants' lessened engagement with the experiment will diminish both its external validity (in light of entrepreneurship's characteristics) as well as its construct validity – in the sense that this lessened engagement will introduce confounding dynamics in the research, thus undermining researchers' ability to correctly infer the theoretical meaning of the causal effects of interest.

A third illustrative tradeoff concerns the particularly high levels of task demands, information overload, time pressures and other practical constraints typically characterizing entrepreneurial pursuits (cf. Baron, 1998; Sutcliffe and Weick, 2008). By definition, experimental methods presuppose a surgical focus on manipulating just a few target factors of interest – while leaving everything else constant or unchanged. This key feature of experiments is what allows them to provide evidence in support of their manipulated factors' causal effects (Colquitt, 2008; Highhouse, 2009; Shadish et al., 2002). Seen in this light, the externally-desirable idea of adding conditions of multiple task demands, information overload, time pressures and/or other constraints to an entrepreneurship experiment would risk inducing other dynamics that could interact with participants' reaction to the manipulated variable(s) of interest, thus lessening the manipulations' construct validity and one's ability to correctly infer the manipulated variables' causal effects. Unless one is specifically investigating the effects of such conditions (e.g., Iselin, 1989), it would thus seem preferable to articulate entrepreneurship experiments such that they exhibit a defensible equilibrium between the ideals of real-life representativeness (implying higher levels of task demands, etc.) and that of a “clean” and “focused” research design (implying lower levels of task demands, etc.).

What unites the above tradeoffs is that they directly concern a balancing act between the ideals of external and construct validity. On the one hand, the ability to derive meaningful generalizations from an entrepreneurship experiment depends on the extent to which its manipulation(s), material and procedures are sufficiently representative of the most pertinent characteristics of the entrepreneurial phenomenon of interest. On the other hand, the ability to derive meaningful theoretical insights from an experiment depends on the extent to which its focal manipulations are sufficiently aligned with relevant variations of the theoretical construct(s) of interest – and are considered as such by participants, in the experiment's particular context and circumstances. Without this balancing act, entrepreneurship experiments can only offer suboptimal evidence in support of the causal effects they attempt to demonstrate. In order to assess extant practices and eventually formulate pertinent recommendations to help entrepreneurship researchers achieve this delicate balancing act, we briefly summarize each set of validity issues and relevant strategies.

2.2. The external validity of entrepreneurship experiments

External validity concerns the extent to which an experiment's findings can be generalized across populations, settings, manipulations and outcomes (Highhouse, 2009: p. 555; Shadish et al., 2002: p. 83). Yet as noted by Crano and Brewer (2002), “external validity is not a single construct but represents a whole set of questions about generalizability, each with somewhat different implications for the interpretation and extension of research findings (p. 108; citing also Brewer, 2000).” These authors thus distinguish

³ Naturally, we would relax this position for studies specifically interested in examining the effects of various levels of uncertainty/stage of advancement on participants' ideation and/or evaluation efforts (cf. Gustafsson, 2006; Holm et al., 2013; Koudstaal et al., 2015; McKelvie et al., 2011).

between the three subordinate issues of an experiment's *robustness* (can a study's findings be replicated?), *ecological validity* (is an experiment representative of what happens in everyday life?) and *relevance* (do the findings of an experiment matter in practical terms?). For their part, Wilson, Aronson and Carlsmith (2010) unpack the question of external validity by drawing attention to three distinct sets of considerations related to a study's realism. *Experimental realism* concerns the extent to which the data collection procedures and materials are “involving to the participants, if they are forced to take it seriously, if it has impact on them (p. 56).” *Mundane realism* refers “to the extent to which events occurring in the research setting are likely to occur in the normal course of the participants' lives, that is, in the real world (ibidem).” Lastly, *psychological realism* (building on Aronson et al., 1994) concerns “the extent to which the psychological processes that occur in an experiment are the same as psychological processes that occur in everyday life (p. 57).” Consistent with the different nuances they bring forth, each subdimension of external validity calls for particular validation strategies.

2.2.1. Validation strategies for establishing external validity

Common strategies for augmenting an experiment's *ecological validity* and *mundane realism* typically consist of systematic efforts to model a study's procedures, task and material as closely as possible to the real-life situation(s) one attempts to model (Grégoire et al., 2010). Such efforts can also include pre-tests conducted before a focal experiment with relevant practitioners and/or field experts to examine whether the material and/or procedures “make practical sense” and have “face validity” (see Perdue and Summers, 1986).

In similar fashion, common strategies for augmenting a study's *experimental realism* (Wilson et al., 2010) include conducting pilot-tests of the entire data collection procedures (prior to launching the main study) to examine if these pose particular difficulties and/or misunderstandings (Perdue and Summers, 1986). Researchers can also mobilize various post-experimental checks to assess participants' levels of attention, effort and/or engagement with the data collection task and/or material.

Establishing the *robustness* of an experiment's findings (Crano and Brewer, 2002) calls for replicating it with various sample frames, data collection procedures and/or research tasks and materials (see Bergh et al., 2017 and Bettis et al., 2016 for related discussions).

For their part, issues of *psychological realism* (Wilson et al., 2010) often arise with experiments that mobilize abstract or decontextualized research tasks and material, yet are nevertheless assumed to tap into psychological or behavioral dynamics that are the same as those that occur in real life. Typical strategies to establish psychological realism tend to be rhetorical – such as drawing from relevant studies to establish the assumptions undergirding an experiment's procedures and/or explaining how and why the abstract/decontextualized procedures effectively serve to “magnify” relevant thought processes, behaviors or social interactions that are otherwise very common and taken-for-granted in everyday life.

Lastly, challenges of *relevance* (Crano and Brewer, 2002) are typically handled through rhetorical arguments that a study's findings matter for practice. In the same spirit, one can also mobilize effect-size measures (Olejnik and Algina, 2003; Pierce et al., 2004) and other translations of a study's statistical results to offer convincing arguments of its focal relationships' meaningfulness (cf. Aguinis et al., 2018; Edwards and Berry, 2010; and Goldfarb and King, 2016 for related observations).

2.2.2. From generalizability to representativeness

In spite of their importance for establishing various dimensions of an experiment's external validity, the fine-grained strategies above remain subservient to broader pragmatic concerns. In an applied field of social-science research like entrepreneurship, researchers, educators and practitioners are typically *less* preoccupied with the ability to generalize a study's findings across each and every possible kind of individuals and contexts (as one would in economics or general psychology), and more with the ability to generalize a study's findings *to the specific people and circumstances of interest to entrepreneurship research*. This difference directly echoes Highhouse (2009)'s distinctions between ecological validity and external validity: while external validity concerns the ability to generalize a study's findings across *different* contexts and populations (in reality), ecological validity concerns the ability to generalize these findings to a *specific* context or population.

Since entrepreneurship research concerns the specific circumstances and dynamics of entrepreneurial action (cf. Davidsson, 2016; Gartner, 2001; Shane and Venkataraman, 2000), the validity challenges of entrepreneurship experiments are arguably less centered on generalizing evidence for the causal effects of entrepreneurship-fostering dynamics *beyond* entrepreneurship people and pursuits than they may be with generalizing these findings to the very actors, circumstances, and processes that make up entrepreneurship-relevant pursuits in real-life. This is not to say that broad generalizations are irrelevant. But in practice, entrepreneurship scholars, journal reviewers and editors tend to be more preoccupied with ecological validity questions of *sampling and procedural representativeness*. We discuss each in turn.

2.2.3. Sampling representativeness

The first representativeness challenge concerns the similarity of an experiment's sample frame and actual participants with respect to the individuals, groups, firms and/or larger social entities involved in the real-life phenomenon one is studying. The higher this similarity, the more confident one can be that an experiment's results are representative of the real-life unfolding of the phenomenon of interest – and therefore, that these findings generalize to other relevant populations specifically concerned with the phenomenon of interest. To foster sampling representativeness, it is thus common to conduct an experiment with participants whose demographics, experiences, knowledge and other motivations are similar to that of the typical populations involved in the phenomenon of interest.

This issue of sampling representativeness is perhaps best illustrated by recurring debates about the use of student samples in entrepreneurship research (Kraus et al. 2016: p. 962; Urbig, 2016) – and more recently, with issues about using sample frames from Amazon™'s Mechanical Turk® (MTurk®) and other crowdsourcing panels (Keith and Harms, 2016; Woo et al., 2015).

With respect to student samples, Hsu et al., (2017a: pp. 385–6) list three conditions under which such samples may be particularly appropriate (but see also Bello et al., 2009; Camerer and Fehr, 2006; Gordon et al., 1987; and Stevens, 2011). The first condition is “when students are or resemble the population of interest (Hsu et al., 2017a: p. 385)” – such as when examining the outcomes of new education programs (Sánchez, 2013; Souitaris et al., 2007; Volery et al., 2013), the formation of entrepreneurial intentions (Fitzsimmons and Douglas, 2011; Gupta et al., 2008) or the behaviors of first-time founders with minimum levels of experience (McGee et al., 2009). The second condition is “when the manipulation is likely confounded by the professional experience of the participants (Hsu et al., 2017a: p. 385)” – such as when an experiment involves the manipulation of a factor that could be strongly confounded, influenced or even overruled by the occupational biases and real-world experiences of representative individuals from the target population (for instance, “a manipulation designed to alter entrepreneurial self-efficacy is likely to be confounded with the prior success and failure experiences of entrepreneurs (Hsu et al., 2017a: p. 385)).” The third condition is “when the relationships under investigation are grounded in a broad theory (Hsu et al., 2017a: p. 385)” that has been shown to apply under various populations and contexts (as in the case of Kahneman and Tversky's (1979) prospect theory, see Hsu et al., 2017b).

With respect to using samples from Amazon™'s MTurk® and/or other crowdsourcing platforms, a key sampling representativeness challenge arises from the limited information typically offered about the respondents. Although such panels might offer relatively easy access to a broad diversity of participants with different backgrounds and characteristics (Clifford et al., 2015; Keith and Harms, 2016) and readily allow for specifying inclusion/exclusion criteria ahead of time, some studies have indicated that participants to crowdsourcing panels tend to be somewhat younger, more educated, less employed, less religious and more liberal than the general population – at least in the United States where most M-Turk® workers are said to reside (see Berinsky et al., 2012; Shapiro et al., 2013). More pointedly, some authors have remarked that crowdsourcing platforms' typical lack of transparency with respect to the recruitment and identities of their workers makes it difficult to establish the particular composition of the obtained samples – even when one requests particular sampling frame criteria (Cheung et al., 2017). Such transparency concerns might pose even more validity problems when trying to sample populations with very specific occupations – as is the case with entrepreneurs – let alone in the face of mounting worries about the use of automated task-completion software (see Dreyfuss, 2018).

As the two cases of students and crowdsourcing samples illustrate, the relevant external validity challenges associated with sample representativeness pertains to the extent to which one can legitimately defend that the experimental findings obtained from a sample should generalize – in principle – to what would otherwise be observed with representative members of the target population most directly concerned with the phenomenon of interest. In practice, thus, a common strategy for augmenting sampling representativeness consists of conducting an experiment with participants whose abilities, demographic characteristics, prior experiences, knowledge, personal motivations, affective and cognitive dispositions, level of engagement and other personal circumstances are as similar as possible to those of the target group of real-life people most concerned with the study's phenomenon of interest.

2.2.4. Procedural representativeness

For its part, the second representativeness challenge concerns the validity of an experiment's data collection procedures, research task and material with respect to the particular phenomenon one is investigating. The challenge is to design experiments that are as realistic as possible with respect to the salient characteristics of entrepreneurial action. It thus concerns the aforementioned notion of *ecological validity* (Crano and Brewer, 2002), but also that of relevant nuances of experimental, mundane and psychological realism/relevance (Wilson et al., 2010). As a result, strategies for augmenting procedural representativeness typically consist of explicit efforts to design an experiment's procedures, tasks and materials such that they closely reflect the real-life phenomenon of interest – and in ways that closely align with the manner how real-life agents experience this phenomenon.

To study the effects of entrepreneurs' displayed passion on investment decisions, for instance, Chen et al. (2009, Study 1) created a series of video scripts where the content being communicated to participants was directly based on real-life business plans from a prior entrepreneurship competition. In turn, they recruited a professional actor to record these scripts while conveying varying degrees of passion. From a representativeness standpoint, these authors' approach has the dual advantage of stimuli that are representative of real-life (business plans *and* behavioral displays of passion) but also, of mobilizing a research task (evaluating a pitch that includes both audio and visual stimuli) that closely resembles what investors typically encounter in their day-to-day screening activities. Taken together, these research material and research procedures work hand-in-hand to increase the study's procedural representativeness (see also Clarke et al. (in press) for a similar approach).

2.3. The construct validity of entrepreneurship experiments

Having summarized the main issues and strategies relevant for establishing the external validity of entrepreneurship research experiments, we turn our attention to construct validity. From a definitional standpoint, the notion of construct validity concerns the ability to “(make) inferences from the sampling particulars of a study to the higher-order constructs they represent (Shadish et al. 2002: p. 65).” To the extent that we want to establish the veracity of the theoretical developments advanced in an experimental study, “we need to know if our treatments have been perceived or interpreted as we intended them to be (Crano & Brewer 2002: 82–3).” Construct validity thus concerns the operationalization(s) of an experiment's focal variables – the meaning of its manipulations so-to-speak (Wilson et al., 2010).

2.3.1. Validation strategies for establishing construct validity

Strategies to increase construct validity start from a sound command of the relevant theoretical and practical considerations surrounding an experiment's focal variable(s). By definition, the power of experimental methods for establishing causality not only

rests on designing manipulations that reflect variations of theoretical interest, but also on designing manipulations that a study's participants effectively perceive, interpret and/or process in a theory-consistent manner. When prior experiments involving the same construct(s) exist, it becomes possible to import and/or build on these experiments' manipulations. When no such exemplars can be found, however, one must embark on a creative journey to articulate a pool of defensible manipulations (in similar fashion to the strategies advocated for developing new survey instruments; cf. DeVellis, 2016; Furr, 2011; Hinkin, 1998).

Yet whether starting from scratch or building on prior exemplars, establishing construct validity ultimately rests on deliberate empirical efforts to *measure* the extent to which an experiment's proposed manipulations both align with theory, and are treated in a theory-consistent manner by participants (Perdue and Summers, 1986). These tests can take many shapes, they can be conducted with the same or different participants, using the same or different forms of data collection (e.g., qualitative vs. quantitative, with implicit or explicit measures). By and large, however, methodologists distinguish two primary forms of tests for assessing a manipulation's construct validity (cf. Crano & Brewer 2002: 79–85; Perdue and Summers, 1986; Wilson et al. 2010: 65–77): *ex-ante manipulation tests* typically conducted with different participants than the focal experiment, and *post hoc manipulation checks* conducted with the same participants.

2.3.2. Ex-ante manipulation tests

Conducted *before* the main data collection, *ex-ante manipulation tests* arguably promise the highest potential value to experimenters (Aronson and Carlsmith, 1968), in large part because they allow for adjusting/modifying the manipulations before experimenters engage in their main data collection events. In practice, conducting such tests often calls for experimenters to develop parallel studies from their main experiment and/or to mobilize different data collection techniques. For instance, one could conduct a pre-test with academic experts to examine the extent to which the proposed manipulations align with the definition of theoretically-relevant variations (see Hinkin and Tracey, 1999 for a parallel application of such techniques in the context of developing new survey questions). One could also mobilize interviews and/or open-ended survey questions with field experts and/or target participants to check whether the proposed manipulations have face validity. In their efforts to design an experiment on entrepreneurs' decisions *not* to pursue an opportunity, for instance, Wood et al. (2017) conducted an online survey with entrepreneurs specifically asking them about their experience vis-à-vis opportunities they had chosen not to pursue.

In addition to such explorative pre-tests conducted to help design a study's manipulation, different tests can be mobilized to investigate whether the participants to an experiment perceive, interpret and treat the proposed manipulations in a theory-consistent manner. These tests are particularly important to establish that the causal claims derived from an experiment are meaningfully-valid from the participants' standpoint. In their experimental manipulation of social approval in video pitches, for instance, Mueller et al. (2018) not only needed to demonstrate that the cues they had introduced to their crowdfunding video pitches effectively corresponded to very distinct types of social approval (low vs high); they also needed to ensure that the manipulation solely influenced participants' social approval perceptions – and not their perceptions of creativity, novelty or usefulness. To this end, these authors conducted an online exercise where they randomly assigned either a *low* or *high* social-approval version of their crowdfunding idea to 93 MTurk® participants, and simply asked them to rate the idea among the four different dimensions of social approval, creativity, novelty, and usefulness. Assuming that responses from these crowdsourcing participants are representative of the general population's interpretation and understanding of such stimuli (but see above), these authors showed that their manipulation changed participants' social approval perceptions but *not* their ratings of the idea's creativity, novelty, or usefulness. These results thus validated that the authors' manipulations effectively varied the construct of interest – and only that construct. In short, the idea of *ex-ante* manipulation tests is to mobilize slightly different designs, techniques and procedures (from the main experiment) to document that the intended manipulations translate into theory-consistent changes, perceived and mobilized as such by representative participants.

2.3.3. Post hoc manipulation checks

For their part, *post hoc manipulation checks* are typically mobilized to serve either of three related purposes. In some experiments, it becomes paramount to establish that participants were *not* aware of the study's goal or manipulation – as this would potentially bias the results (Crano & Brewer 2002: 82–84; Perdue and Summers, 1986; Wilson et al. 2010: 68). This is notably the case when a study concerns socially-desirable issues, and/or when a manipulation could potentially induce some demand characteristics (see Fisher, 1993; Lautenschlager and Flaherty, 1990). Although this cannot establish construct validity in and of itself, such tests help establish that an experimental manipulation was not so immediately transparent that participants' responses have more to do with what they think experimenters want to find than with what they would report otherwise.

Manipulation checks can also be conducted *after* the main outcome measures have been collected to assess the extent to which participants' naïve interpretation of the study's procedures and material effectively matches the manipulations' theoretically-intended meaning. In their study of the impact of investors' ethical reputation, for instance, Drover et al. (2014: Study 2) presented entrepreneurs with a short scenario describing a particular venture capital (VC) firm that wanted to invest in the participants' ventures. Embedded in the scenarios were a few sentences manipulating the VC firm's ethical reputation – in their case, the extent to which the VC firm i) tended to favor premature exits in order to rapidly cash-in their gains at the expense of the entrepreneurial venture's growth potential (unethical), ii) tended to wait until higher growth and returns could be achieved (ethical), or iii) varied its preferences depending on the circumstances (neutral). Once they had obtained participants' answers to the experiment's outcome variable (participants' willingness to partner with the VCs), the authors conducted a manipulation check where they simply asked participants to rate the ethical reputation of the VC firm (on a 1-to-7 Likert scale). Consistent with their manipulation's intent, the authors reported statistically-significant evidence that the mean rating reported by participants who had received the ethical scenario

was higher than that of participants who had received the unethical scenario. This helped establish that indeed, participants had perceived and processed the manipulated scenarios in the manner the authors intended.

Lastly, some experiments mobilize post hoc manipulation checks to examine the extent to which a manipulation had lasting effect (see [Dvir et al., 2002](#)) or as a way to corroborate participants' responses to prior manipulations. In a typical conjoint or other decision-making study where one seeks to uncover participants' in-use/implicit decision policies, for instance, some investigators follow the experiment with a series of questions asking participants to rate the importance they attach to different criteria (e.g., [Shepherd, 1999a](#); [McKelvie et al., 2011](#)) – or simply ask them to describe how they answered the exercise through post-experiment's interviews (e.g., [DeTienne et al., 2008](#)). Experimenters can then examine whether participants' experimental responses correlate with their post hoc answers, thereby providing evidence that participants' *in-use* decision policies (i.e., those derived from analyzing the pattern of responses from the experiment) effectively reflect their *espoused* decision policies (i.e., what they openly report as more or less important to them). In their experimental study of entrepreneurs' decision *not* to pursue opportunities, [Wood et al. \(2017\)](#) used a similar post-experiment series of open-ended question to ask participants to explain the rationales for their action and inaction decisions. By content-analyzing these qualitative answers, these authors were able to provide data about their study's realism – and to confirm the central role of doubts in their findings (see pp. 119–121).

Regardless of whether they are conducted before or after a focal experiment, and with different or the same participants, the key objective of manipulation tests and checks remains to provide empirical evidence that an experiment's focal manipulations correspond to meaningful variations of the theoretical construct(s) of interest – and induce theoretically-consistent effects in an experiment's participants. Indeed, both elements are necessary to establish the construct validity of an experiment's findings.

3. Method

To show that addressing the validity tradeoffs inherent to entrepreneurship research experiments is not a trivial task, we conduct a structured literature analysis (cf. [Jones and Gatrell, 2014](#); [Tranfield et al., 2003](#)) that documents the mobilization of external and construct validation strategies in a relevant corpus of entrepreneurship research experiments. In turn, we use the findings to formulate a pertinent set of recommendations to facilitate the sound mobilization of experiments in entrepreneurship research.

3.1. Search procedures and sampling frames

We describe in [Appendix A](#) (at the end of this article) the search procedures and inclusion/exclusion criteria we mobilized to identify the relevant studies forming two comprehensive samples of entrepreneurship research experiments we subsequently analyzed: 1. A first sample of 100 research articles published in seven entrepreneurship-specific journals listed in Thompson Reuter's Journal Citation Reports® (JCR) – Social Sciences Edition, 2016, and 2. A comparative sample of 44 research articles focused on entrepreneurship-relevant topics and published in JCR-listed journals targeting audiences in applied psychology, behavioral economics, and the managerial and organizational sciences. We use this dual-sample strategy as a robustness check allowing us to examine whether validation practices vary between journals targeting different academic audiences.⁴ Supplemental Online Appendix B provides the full references to all the papers forming the two samples.

3.2. Analytical schemes and coding procedures

In order to assess the validation efforts of extant entrepreneurship research mobilizing experimental methods, we analyzed the corpus of identified articles in light of two sets of coding schemes specifically formulated to tap into the particular issues of external and construct validity. To examine entrepreneurship research experiments' efforts to tackle issues of external validity (and its underlying sub-dimensions), we carefully examined each article to document its mobilization of relevant strategies, using the definitions and indicators summarized in [Table 1](#) (Panel I). In similar fashion, we documented whether each article mobilized one or more of the construct validity strategies listed in [Table 1](#) (Panel II).

Because the indicators for these coding dimensions tend to be expressed directly in the text of a research article, we compiled the data for our analyses by carefully reading the methods and results section of each article in our samples – noting in passing any effort to mobilize the external and construct validity strategies listed in [Table 1](#). To help foster the transparency of our coding but also the rapid identification of the source data (so as to more easily discuss our coding results among ourselves), we documented whether a particular validation strategy was mobilized or not in each article (using a 1,0 dummy-code scheme), as well as the page number(s) where the relevant information could be found (and some comments, when necessary). To reinforce these procedures and help ensure that we did not overlook relevant efforts, we took advantage of the search capabilities afforded with electronic copies of the articles (generally in Adobe®'s Acrobat Reader™'s *.pdf® format) to systematically search for relevant instances of terms like *valid**, *manip**,

⁴ More specifically, this robustness comparison allows us to investigate the extent to which generalist and specialist journals might enforce different validation standards – the more so considering the relative centrality (to these journals) of the validity challenges posed by entrepreneurship's most salient and comingled characteristics. This empirical question is also relevant not only in light of prior criticisms of entrepreneurship research's levels of methodological rigor (e.g., [Crook et al., 2010](#)), but also because it provides authors with relevant observations about current practices. By the same token, this comparison provides journal editors with relevant benchmarks – and with relevant observations to encourage the mobilization of particular validation strategies (see [Cuervo-Cazurra et al., 2016](#)).

Table 1

Coding scheme for documenting validation strategies.

Panel I: External validity (and relevant sub-dimensions.	Definition	Relevant indicators
Ecological validity/mundane realism	The extent to which an experiment reflects the real-life context/phenomenon it means to model or represent (Crano and Brewer, 2002; Wilson et al., 2010)	<ul style="list-style-type: none"> ● Pre-test of an experiment's research material and/or procedures with relevant field experts who could then attest that the material and/or procedures “makes practical sense” or has “face validity” (see Perdue and Summers, 1986)
Experimental realism	The extent to which an experiment's data collection procedures and materials are “involving to the participants, if they are forced to take it seriously (and/or) if it has impact on them (Wilson et al. 2010: 56).”	<ul style="list-style-type: none"> ● Pilot-test of an experiment's entire data collection procedures to assess if any of it posed particular difficulties (Perdue and Summers, 1986) ● Post-experiment's check to assess participants' overall levels of attention, effort and/or engagement with the data collection task and/or material.
Robustness	The extent to which an experiment's findings are found to applicability to other data collection procedures and settings (Crano and Brewer, 2002)	<ul style="list-style-type: none"> ● Explicit effort to replicate an experiment's focal findings through multiple experiments (including with different sample frames) ● Explicit effort to replicate an experiment's focal findings through another data collection effort using one or more different method(s). ● Use of the manipulation(s) not to test a focal hypothesis, but to assess the stability of another relationship across different conditions. ● Reporting of test-retest reliability results, implying multiple dependent variable observations for the same manipulated stimuli
Relevance	The extent to which an experiment's findings “matter” in a practical sense (Crano and Brewer, 2002)	<ul style="list-style-type: none"> ● Reporting of effect sizes ● Comments about the magnitude of observed relationships in terms of their underlying variables and metrics
Other		<ul style="list-style-type: none"> ● Any other strategy or procedure argued to help establish a study's external validity
Panel II: Construct validity	The extent to which the manipulations of a target construct (in terms of their concrete articulations) corresponds to meaningful variations of the underlying theoretical construct	<ul style="list-style-type: none"> ● Separate construct validity study conducted with participants from a different but pertinent sample frame, and focusing on assessing the alignment/coherence between the manipulations' theoretically-intended effects and their effective impact on participants; ● Post-experiment manipulation check conducted with participants from the focal study, and using different data/task(s) to assess participants' awareness of a study's manipulation; ● Post-experiment manipulation check conducted with participants from the focal study, and using different task(s) to assess the manipulation's effectiveness in light of its theoretically-intended effects; ● Lastly, we also considered whether the author(s) of an article argued for the mobilization of any other strategy or procedures meant to augment and/or establish their work's construct validity.

check, pilot, pretest, pre-test and/or effect size. We report in [Appendix C](#) (at the end of this article) the procedures we implemented to ensure the consistency and reliability of our coding.

3.2.1. Supplemental analyses

While the raw data for documenting the mobilization of the external validity strategies listed in [Table 1](#) can be found directly in the text of our samples' articles, we supplemented our inventory with two subjective evaluations we compiled after reading each article. Mobilizing such evaluations allows us to examine whether the possible absence of formal validation strategy might be compensated by research designs that already exhibit such high levels of external validity that formal checks and other empirical tests would seem redundant, superfluous or even counterproductive.

The first supplemental assessment focused on each study's **sample frame representativeness**, which we documented by coding the extent to which the participants of an experiment (sampling frame) exhibit similar demographic characteristics and levels of familiarity, knowledge, motivations and/or personal relevance than the typical agents involved in the topic/phenomenon one attempts to study with this experiment (answered on a five-point scale anchored 1 = very little; 3 = moderate; and 5 = very much so). Conducting this supplemental assessment allowed us to evaluate the methodological adequacy of collecting data from a particular sample frame of respondents/units, given the particular phenomenon modeled with an entrepreneurship research experiment. Although our measure consists of an overall subjective rating, we based this rating on a series of similarity indicators focusing on whether an experiment's participants and the real-life agents they are meant to represent share: i) similar demographic characteristics (age, gender, ethnicity, social-economic status, etc.); ii) similar levels general knowledge and familiarity with the phenomenon of interest, as obtained

though general education and/or experience; iii) similar levels of specific ability with phenomenon of interest, as obtained through specific training and/or experience (i.e., participants have done this specifically before); and/or iv) similar real-life circumstances in terms of ambitions, goals and motivations, career stage, development of the business and organization, etc.

For the second supplemental assessment, we assessed each study's *procedural representativeness* by coding the extent to which the research's instructions, task(s), outcome measure(s), material and other stimuli are “similar” to the real-life circumstances pertinent to the topic/phenomena of interest of the study (answered on a five-point scale anchored 1 = very little; 3 = moderate; and 5 = very much so). In many ways, this rating scheme taps into the “richness” of an experiment's procedures and material vis-à-vis the real-life circumstances it means to reflect. As with the above rating scheme, we based our coding on a series of similarity indicators focusing on the extent to which an experiment's procedures and the circumstances they attempt to represent share: i) similar instructions (i.e., what participants are told to do); ii) similar tasks and activities (i.e., what participants are asked to do in the experiment); iii) similar outcomes (i.e., the measure of what people are doing); iv) similar material/input stimuli (both in terms of content and format); and/or v) similar research design (i.e., single or repeated nature of the procedures).

Like with the primary content analyses above, we report in [Appendix C](#) the procedures we implemented to ensure the consistency and reliability of our coding for sample frame and procedural representativeness.

4. Results

4.1. The mobilization of external validation strategies

[Table 2](#) displays the mobilization of external validation strategies in our two samples of entrepreneurship research experiments. To examine the possibility of temporal variations in the mobilization of such strategies, we broke down the overall results (reported in the first column for each sample) into three periods of unequal length but similar number of articles.

With respect to the *ecological validity* and *mundane realism* of entrepreneurship experiments ([Table 2](#), line 1), the use of pre-tests to establish the representativeness of a study's material is the most frequently used external validation strategy: 43.00% of experiments published in entrepreneurship journals and 31.82% of experiments published in our sampling frame's generalist journals report having conducted pre-tests of their research material. Curiously, however, both samples exhibit a downward temporal trend, with pre-tests being reported less frequently in studies published in 2015–17 (with figures of 31.25% and 18.20%) than in studies published between 2010 and 2014 (40.00% and 41.70%), and even more so by comparison to studies published prior to 2010 (55.26% and 50.00%).

For their part, issues of *experimental realism* appear less frequently addressed ([Table 2](#), lines 2–3). Indeed, the conduct of pilot-tests examining the entire data collection procedures is typically reported in less than a quarter of the articles (with overall figures of 23.00% and 18.18% for the two samples). One notices some temporal variations though, with an apparent upward progression in entrepreneurship journals (from 18.42% to 31.25%) and a downward progression in generalist journals (from 30.00% to 9.10%). In similar fashion, post-experiment checks to assess participants' attention, efforts and/or engagement are typically reported in < 10% of entrepreneurship experiments.

The mobilization of strategies for establishing the *robustness* of experimental findings exhibits more variations ([Table 2](#), lines 4–7). The reporting of multiple experiments within the same article typically occurs in about 10–20% of the corpus, with averages of 15.00% and 11.36% for the two samples, and an increasing trend in studies published in entrepreneurship journals (up to 21.88% in 2015–17). By contrast, the reporting of parallel studies using different data collection techniques within the same article appears more frequent in entrepreneurship research experiments published in generalist journals (with an overall figure of 25.00%) than those published in entrepreneurship journals (8.00%) – but this strategy's mobilization appears to greatly vary in time, with both samples exhibiting a fair range between their high and low values (16.66% vs. 3.13% in Sample 1, 40.00% and 8.33% in Sample 2). For its part, the strategy of using manipulations to assess the stability of a relationship across conditions is rarely used in entrepreneurship journals (with only three studies doing this in 2015–17), but it appears more common in generalist journals (appearing in 9.09% of the articles in this sample). Interestingly, anecdotal observations suggest that this kind of approach might be more typical of experiments relying on behavioral economics tasks such as games, lotteries and choices to examine the implications of individual-level variations in risk-related attitudes and preferences. Lastly, the use of test-retest reliability assessments occurs in about 27–28% of articles (for both samples). But robustness analyses (see below) indicate that this observation is primarily driven by the decision experiments: few other studies conduct such tests.

With respect to establishing the *relevance* of experimental findings ([Table 2](#), line 8), about a third of the experiments published in entrepreneurship journals (31.00%) offer some comments or measures specifically discussing the magnitude of the observed effects in some form of meaningful terms (e.g., effect sizes, % changes, increases in terms of original metrics, etc.). Interestingly, we do not observe much temporal variation around this average, with figures varying from 26.32% for articles published prior to 2010 to figures of 33.33% and 34.38% for 2010–14 and 2015–17. By contrast, commenting on the magnitude of one's statistical findings appear more prevalently among studies published in generalist journals, with an average of 38.64% and a high of 54.50% for studies published since 2015.

Supplementing the above findings, lines 9 and 10 report the overall assessments we made of the studies' sample frames and procedural representativeness. On the one hand, the figures for *sample frame representativeness* (3.90 and 3.56 for the two samples) suggest that the “average” experiment conducted in entrepreneurship research is typically conducted with samples of participants that share relatively high levels of similarity with representative members of the relevant target population most concerned with a study's phenomenon of interest. The being said, the slight downward temporal trend (in both samples) and more importantly, the

Table 2
The mobilization of external validation strategies in entrepreneurship experiments.

Sub-dimensions	Strategies	Number (%) – Sample 1 JCR®-listed entrepreneurship journals				Number (%) – Sample 2 JCR®-listed social-science journals			
		Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)
		n = 100	n = 38	n = 30	n = 32	n = 44	n = 10	n = 12	n = 22
Ecological validity/ mundane realism Experimental realism	1. Pre-test of an experiment's research material to establish that it is representative	43 (43.00)	21 (55.26)	12 (40.00)	10 (31.25)	14 (31.82)	5 (50.00)	5 (41.70)	4 (18.20)
	2. Pilot-test of an experiment's entire data collection procedures	23 (23.00)	7 (18.42)	6 (20.00)	10 (31.25)	8 (18.18)	3 (30.00)	3 (25.00)	2 (9.10)
	3. Post-experiment's checks to assess participants' attention, effort and/or engagement	7 (7.00)	3 (7.89)	1 (3.33)	3 (9.38)	2 (4.54)	0	1 (8.30)	1 (4.50)
Robustness	4. Multiple experiments within single paper	15 (15.00)	5 (13.16)	3 (10.00)	7 (21.88)	5 (11.36)	0	2 (16.67)	3 (13.64)
	5. Parallel study using other data collection methods	8 (8.00)	2 (5.26)	5 (16.66)	1 (3.13)	11 (25.00)	4 (40.00)	1 (8.33)	6 (27.27)
	6. Use of manipulation(s) to assess the stability of another relationship across different conditions	3 (3.00)	0	0	3 (9.38)	4 (9.09)	1 (10.00)	2 (16.67)	1 (4.50)
Relevance	7. Test-retest reliability assessment	28 (28.00)	11 (26.32)	9 (33.33)	8 (34.38)	12 (27.27)	5 (50.00)	4 (3.33)	3 (13.60)
	8. Reporting of effect sizes and/or comments about the magnitude of observed relationships in terms of their underlying variables and metrics	31 (31.00)	10 (26.32)	10 (33.33)	11 (34.38)	17 (38.64)	3 (30.00)	2 (16.70)	12 (54.50)
Sub-dimensions	Strategies	Average ratings (5-point scale) (standard deviation)				Average ratings (5-point scale) (standard deviation)			
		Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)
		n = 100	n = 38	n = 30	n = 32	n = 44	n = 10	n = 12	n = 22
Overall assessments	9. Sample frame representativeness	3.90 (1.18)	4.30 (0.98)	3.60 (1.25)	3.70 (1.24)	3.56 (1.40)	3.60 (1.32)	3.95 (1.07)	3.33 (1.59)
	10. Procedural representativeness	3.11 (1.19)	3.43 (1.15)	2.72 (1.03)	3.09 (1.31)	3.25 (1.18)	2.80 (1.01)	3.06 (1.16)	3.56 (1.22)

relatively large and increasing standard deviations (> 1 on a 5-pt scale) suggest that studies with more questionable sampling frames continue to be published alongside otherwise excellent studies. Reinforcing this observation, the figures for *procedural representativeness* (3.11 and 3.25 on average for the two samples, again with standard deviations larger than one) suggest that the “average” experiment published in entrepreneurship research does not typically mobilize research material, tasks and procedures that already exhibit such high-levels of similarity with respect to the real-life phenomenon it attempts to model that it can reasonably forego the mobilization of other external validation strategy.

All in all, the above results indicate that while many entrepreneurship research experiments mobilize strategies relevant to establishing their findings' external validity, the mobilization of such strategies does not appear systematic – including in papers published in recent years or in journals targeting different academic audiences. Fazed with this result, we conducted a series of robustness checks to investigate if this overall pattern also characterized sub-samples of experiments that are more frequently conducted in entrepreneurship research – namely decision experiments, information manipulations and quasi-experiments. We report these additional observations in [Appendix D](#) (at the article's end) yet note that the findings obtained reinforce the above observations.

4.2. The mobilization of construct validation strategies

[Table 3](#) displays the mobilization of construct validation strategies in our two samples of entrepreneurship research experiments. By and large, results suggest that a minority of entrepreneurship research experiments mobilize formal strategies to establish their manipulations' construct validity.

For instance, 13.00% and 15.91% of the experiment studies published in entrepreneurship and generalist journals (respectively) report having conducted a separate construct validation study; across the different time periods, the figure never reaches above 20% (see [Table 3](#), line 1). With average figures of 9.00% and 11.36% for the two samples, similar observations can be made with respect to the conduct of post-manipulation checks regarding participant's understanding of the material (see [Table 3](#), line 2). We did not observe any use of post-experiment checks to assess manipulation awareness (line 3). More positively, 37.00% of the experimental studies published in entrepreneurship journals and 31.82% of those published in generalist journals mobilized some form of post-experiment checks of manipulation effectiveness (line 4). And in both cases, there is evidence of an increasing temporal trend, at least when comparing studies published in 2010–14 with those published in 2015–17 (with increases from 26.67% to 37.50% in Sample 1, and from 16.67% to 40.91% in Sample 2). In spite of these positive observations, however, a majority of entrepreneurship experiments published in either sample does not report the mobilization of *any* strategy to establish their construct validity (line 5). As we did with external validation strategies, we conducted a series of robustness checks to investigate whether the mobilization of construct validation strategies might vary across decision experiments, information manipulations and quasi-experiments. Reported in [Appendix D](#), the results support the above observations.

5. Discussion

The use of experimental methods is gaining momentum in entrepreneurship research. Indeed, the sheer diversity and additional numbers of experimental studies we identified in our literature search reinforce prior observations from [Kraus et al. \(2016\)](#) and [Hsu et al. \(2017\)](#). Considering these methods' advantages for examining causal relationships ([Colquitt, 2008](#); [Grégoire et al., 2010](#); [Highhouse, 2009](#)), integrating experiments alongside other research techniques offers promising avenues for augmenting academic understanding of entrepreneurship ([Davidsson, 2016](#)). Our work makes three contributions to help realizing this potential: 1) we draw attention to the particular validity tradeoffs hindering entrepreneurship research experiments, 2) we empirically document that addressing these is not a trivial task, and 3) we build on these findings to offer guidance on relevant strategies for bolstering the validity of future experiments.

5.1. The peculiar nature of validity tradeoffs in entrepreneurship research experiments

Our work complements the generic observations typically found in a host of research methods monographs by presenting an overview of the critical validity tradeoffs inherent to entrepreneurship research experiments. In this regard, we advanced the notion that because the more salient characteristics of entrepreneurship are comingled with one another, the conduct of methodologically-sound entrepreneurship research experiments requires a balancing act between one's efforts to design experiments that reflect the complex and multifaceted aspects inherent to the entrepreneurship phenomenon of interest, and to design manipulations that are both consistent with theory and processed as such by an experiment's participants. When entrepreneurship research experiments fall short on these fronts, doubts arise about the evidence they offer in support of their theorized causal effects. Addressing these tradeoffs calls for entrepreneurship researchers to mobilize various strategies to strengthen their experiments' external and construct validity. By carefully discussing these entrepreneurship-specific tradeoffs and combining that with an overview of relevant validation strategies, we offer an entrepreneurship-centered synthesis that invites researchers to become leaders in the advancement and application of experimental methods.

5.2. Prior efforts to mobilize formal validation strategies

Our systematic analysis of the literature makes a pertinent empirical contribution: that of documenting the mobilization of external and construct validation strategies in two comprehensive samples of entrepreneurship articles that reported findings from

Table 3
The mobilization of construct validation strategies in entrepreneurship experiments.

	Construct validation strategies				Number (%) – Sample 1 JCR®-listed entrepreneurship journals				Number (%)– Sample 2 JCR®-listed social-science journals			
	Prior to 2010		2010–14 (incl.)		2015–17 (incl.)		Overall		Prior to 2010		2010–14 (incl.)	
	n = 100	n = 38	n = 30	n = 32	n = 32	n = 44	n = 10	n = 12	n = 22	n = 22	n = 22	n = 22
All studies in each sample												
1. Separate construct validity study	13 (13.00)	2 (5.26)	6 (20.00)	5 (15.63)	5 (15.63)	7 (15.91)	1 (10.00)	2 (16.67)	4 (18.18)			
2. Post-manipulation checks of material's understanding	9 (9.00)	1 (2.63)	5 (16.67)	3 (9.38)	3 (9.38)	5 (11.36)	2 (20.00)	0	3 (13.64)			
3. Post-experiment checks of manipulation awareness	0	0	0	0	0	0	0	0	0			
4. Post-experiment checks of manipulation effectiveness	37 (37.00)	17 (44.74)	8 (26.67)	12 (37.50)	12 (37.50)	14 (31.82)	3 (30.00)	2 (16.67)	9 (40.91)			
5. No construct validation strategy	52 (52.00)	20 (52.63)	15 (50.00)	17 (53.13)	17 (53.13)	25 (56.82)	5 (50.00)	9 (75.00)	11 (50.00)			

controlled and quasi-experiments.

Among the most positive elements emerging from our analyses, we noted an upward trend in the mobilization of pilot-tests of an experiment's procedures (at least in entrepreneurship journals), upward trends in the mobilization of multiple experiments within a single study (in entrepreneurship journals) and of parallel studies using other data collection methods (in generalist journals), as well as upward trends in the reporting of effect sizes and/or comments on the magnitude of observed effects (in both generalist and entrepreneurship journals). And although the numbers remain small, we observed emerging trends in the conduct of post-experiment tests to assess participants' attention, effort and/or engagement. All in all, these elements signal healthy attention towards issues of experimental realism, robustness and relevance.

To our surprise, however, our analyses also uncovered downward trends in the mobilization of pre-tests to examine the research material's representativeness (in both types of journals), as well as the continued publication of experimental studies that do not report empirical evidence regarding their focal manipulations' construct validity. This is concerning. Without the use of formal pre-tests of a study's research material and procedures, doubts inevitably arise whether an experiment adequately mirrors the real-life experience of entrepreneurs and other relevant agents involved in the phenomenon of interest – in both the stimuli and circumstances they react to and the tasks they perform. And without formal tests of construct validity, one cannot ascertain whether the participants to an experiment perceived, interpreted and processed the manipulated variables in ways that are consistent with the researchers' theoretical intentions. Yet both situations undermine the evidence offered in support of the causal effects of interest, thereby hindering the field's knowledge accumulation and its reputation of methodological rigor.

In the particular context of the *Journal of Business Venturing's* special issue on “Applying experimental methods to entrepreneurship,” these empirical observations allow for taking stock of prior achievements – and of opportunities that lie ahead for researchers to do more in their mobilization of experimental methods. In turn, these findings set the stage for our work's third contribution. Having observed that establishing the validity of entrepreneurship research experiments is not a trivial task, we propose a series of practical recommendations for helping entrepreneurship researchers navigate the validity tradeoffs inherent to conducting realistic and theoretically meaningful experiments.

5.3. A practical guide for facing the validity tradeoffs of entrepreneurship experiments

The desirable end-goal of experimental research is straightforward: to offer the most valid evidence possible in support of the causal effects of a manipulated factor of interest. Seen in this light, the approach we advocate here is *not* to favor one set of validity considerations against another: it would disserve the field to conduct an experiment characterized with extremely high levels of external validity and low levels of construct validity (and vice-versa). Both forms of validity are important for advancing scientific knowledge. The “trick” is to leverage an array of relevant methodological strategies that work together to maximize an experiment's overall validity.

In this regard, the particular challenges of mobilizing experimental methods in entrepreneurship arise because the phenomena of interest typically include a number of salient characteristics that are comingled with one another (e.g., high levels of radical uncertainty, personal stakes and other constraints, etc.). Faced with these challenges, entrepreneurship researchers have frequently resorted to three experimental approaches, each dealing with a particular set of entrepreneurship-relevant characteristics (see [Section 2.1](#)):

1. *Open-ended scenarios* where outcomes typically remain undetermined (to account for entrepreneurship's high levels of uncertainty, ambivalent signals, incomplete information and temporal dynamics);
2. *Hypothetical exercises* that make it ethically- and pragmatically-feasible to study relevant phenomena, but that are challenged to exhibit the same high levels of personal involvement and affective attachment that characterize real-life decisions and pursuits;
3. *Abstract and/or simplified research designs, tasks and procedures* that present “moderate” levels of task demands, information overload and time pressures to allow for better isolating the causal effects of other manipulated factors of interest.

These approaches have many important merits – notably to articulate an experiment's data collection procedures around some of entrepreneurship's most salient characteristics.

To further augment their experiment's external validity, many entrepreneurship researchers design their studies' material on the basis of real-life cases and considerations (e.g., [Chen et al., 2009](#); [Wood et al., 2017](#)). In their recent study of funder self-efficacy and crowd bias in equity crowdfunding, for instance, [Stevenson et al. \(in press\)](#) designed the particular stimuli their experiments' participants would “react to” on the basis of real-life crowdfunding campaigns. In the same spirit, [Hsu et al. \(2017a\)](#) encourage the increased use of “active” manipulations, that is, of research tasks, procedures and material that use deception and/or simulations to “actively” change participants' self-concepts – that is, their own “attitudes, emotions, efficacies, intentions” and other understandings vis-à-vis circumstances that affect them personally (see pp. 389–392). By actively manipulating experimental variables in ways that are meant to directly matter to participants, in very personal ways, this strategy of active manipulations effectively addresses important criticisms regarding the hypothetical nature of many experiments.

Yet in spite of their many advantages, these design approaches are not sufficient on their own to establish the external and construct validity of an entrepreneurship research experiment. To achieve higher levels of validity, it becomes useful to conduct additional data collection efforts to obtain empirical evidence that will rule out alternative explanations and support of a study's external and construct validity. As we documented above, however, the mobilization of such validation strategies has been far from systematic among the controlled and quasi-experiments previously published in entrepreneurship research.

In order to facilitate future uses of experimental methods in entrepreneurship research and their mobilization for establishing the causality of relevant factors and dynamics, we propose a practical guide of validation considerations and strategies. Grouping these together allows us to systematize recommendations typically found across various methods monographs, thereby offering an overall scheme to help entrepreneurship researchers navigate the validity tradeoffs inherent to conducting realistic and theoretically meaningful experiments. To make our recommendations as actionable as possible, we developed an extensive step-by-step guide of design and assessment strategies relevant to entrepreneurship research experiments. The guide spans the entire research process – from conceiving an experimental study to collecting, analyzing and reporting the results. The guide also includes a number of relevant exemplars from many different studies. We present the entire guide in Supplemental Online Appendix E yet summarize the gist of our recommendations in the form of a checklist that experimenters can use as part of their research development efforts (see Table 4).

Our guide's central thrust is to encourage – and equip – experimental researchers to integrate validity considerations across their entire research effort. Indeed, we structured both the extensive guide (in Supplemental Online Appendix E) and Table 4's summary checklist in ways that “situate” particular validation considerations and strategies along different stages of the research process. In this respect, our approach is *not* to require that entrepreneurship experimenters follow each specific recommendation and conduct each validity test listed in our guide. Rather, our hopes are to equip interested researchers with relevant sets of considerations,

Table 4
Proposed guidelines for validating entrepreneurship research experiments.

Stage of research development	Dimension of validity	Checklist of recommended validation strategies (report relevant arguments/ observations in the manuscript)
1. Overall experimental design	External validity (robustness/ replications)	<input type="checkbox"/> Have you considered replicating the experiment across different samples, research material, task and/or procedures, and/or through parallel data collection efforts?
2. Sampling	External validity (sampling frame representativeness)	<input type="checkbox"/> Are you conducting the experiment with participants that are representative of the real-life population most-concerned with the phenomenon of interest? <input type="checkbox"/> If not, are respondents as closely aligned as possible to the target population? <input type="checkbox"/> Obtain measures about similarities between respondents and target population <input type="checkbox"/> If student sample, ensure that its use meets Hsu et al.'s (2017) three conditions <input type="checkbox"/> If crowdsourcing sample, try to obtain verifiable information about respondents and a set of relevant information to establish representativeness vis-à-vis the target population most concerned with the phenomenon of interest
3. Research task	External validity (ecological validity/ mundane realism)	<input type="checkbox"/> Does your research task resemble target population's real-life tasks/activities?
4. Research design	External validity (ecological validity/ mundane realism)	<input type="checkbox"/> Could you conduct your experimental manipulations in real-life settings? <input type="checkbox"/> Is the research best served by a between- or within-subject design? <input type="checkbox"/> Would multiple cases cause manipulation biases? (→between-subject design) <input type="checkbox"/> Can multiple case biases be controlled for? (→within-subject design) <input type="checkbox"/> If within-subject design: do participants encounter manipulations one at a time? <input type="checkbox"/> Can you build manipulations on prior research manipulating same construct? <input type="checkbox"/> If not, can you creatively derive plausible “real-life” articulations of variable?
5. Manipulation/treatment	Construct validity	<input type="checkbox"/> Consider the possibility of actively manipulating participants' self-concept. <input type="checkbox"/> If possible, develop a “pool” of possible manipulations? Select best ones. <input type="checkbox"/> Pre-test manipulations with academic experts (for theoretical consistency)
6. Research material and stimuli	External validity (ecological validity/ mundane realism)	<input type="checkbox"/> Conduct pre-test(s) of manipulation articulations with experts (academic/ field)
	Construct validity	<input type="checkbox"/> Obtain information-rich observations from participants (qualitative tasks) <input type="checkbox"/> Modify manipulations and material according to pre-test observations <input type="checkbox"/> Conduct manipulation validity study(ies) with representative participants <input type="checkbox"/> Obtain information-rich observations from participants (qualitative task) <input type="checkbox"/> If necessary, revise articulation of manipulations and validate revisions
7. Pilot-Tests	External validity (experimental realism)	<input type="checkbox"/> If within-subject design, rule out spillover effect of subsequent manipulations <input type="checkbox"/> Pilot-test the entire set of data collection procedures for external representativeness but also for absence of procedural distractions, issues and other kinks <input type="checkbox"/> Pilot test with experienced/representative members of target population, as well as with less experienced/neophyte participants
8. Data collection	External validity (experimental realism)	<input type="checkbox"/> Observe test participants' level of effort/motivation <input type="checkbox"/> Collect data on participants' level of effort/motivation (post-experiment checks)
	Construct validity	<input type="checkbox"/> Collect data on participants' self-report levels of familiarity with the task <input type="checkbox"/> Collect data on participants' awareness of manipulation (post-experiment checks) <input type="checkbox"/> Collect data on manipulation(s)' effectiveness (post-experiment checks). <input type="checkbox"/> For within-subject manipulations: consider collecting test-retest <input type="checkbox"/> Report your construct validity findings in manuscript
9. Analyses and Results	Construct validity	<input type="checkbox"/> Report “raw” descriptive results like means and other descriptive statistics
	External validity (relevance)	<input type="checkbox"/> Consider reporting and discussing effect sizes <input type="checkbox"/> Comment about magnitude of observed relationships

exemplars and resources from which they can draw inspiration in their own experimental research efforts. Not all experiments are alike. Different purposes and topics call for making different validity choices and tradeoffs. By drawing attention to validity considerations and strategies that seem particularly relevant at different stages, we encourage entrepreneurship experimenters to use their best creative judgment to tailor their validation efforts to the particular challenges and tradeoffs they face – given their research's particular circumstances, purposes, topics and manipulated variables.

By extension, we imagine that our guide might be useful to introduce doctoral students and junior researchers to the nature of construct and external validity challenges in the particular context of entrepreneurship studies – and of the strategies for facing such challenges. With this in mind, we strove to include several exemplars in Online Appendix E that illustrate the design, use and relevance of different validation strategies.

In a similar vein, reviewers and journal editors could mobilize Table 4's checklist in their assessment of manuscripts reporting the findings of one or more experiments. Though we readily see value in such uses, we guard reviewers and editors against a simplistic use of the checklist. As we wrote above, our intent is *not* to ask researchers to conduct every possible test, but rather to draw attention to what can be done, how, and for what validation purpose. In this respect, we stress the importance of integrating validity considerations from the very beginning of one's research efforts. This can not only reinforce the overall rigor of a study, it can prevent the unfortunate situation of having a manuscript returned from a journal because reviewers find too much discrepancies between the experiment's material, procedures and/or participants and the real-life phenomenon of interest – or because they have doubts that participants perceived, interpreted and/or processed the manipulation(s) in the manner that the authors theoretically intended. Seen in this light, we offer that our guide and checklist can provide authors, reviewers and journal editors with an overall set of validity considerations, strategies and practices around which to discuss the relative merits of submitted manuscripts – and for identifying avenues for moving forward.

6. Conclusion

Given that addressing validity tradeoffs is at the heart of conducting experiments in entrepreneurship research, we hope that experimenters will find our synthesis of validation strategies and practical recommendations useful for their efforts to design, conduct, and evaluate experiments that advance understanding of entrepreneurship research. Though we readily acknowledge that not every experiment is subject to each validity concern outlined in this paper (and that no study can possibly rule out every potential validity threat), we vigorously encourage experimenters to devote careful attention to establishing the external and construct validity of their experimental manipulations, material and procedures. By doing so, entrepreneurship researchers mobilizing experimental methods will be better equipped to successfully navigate the thorny methodological tradeoffs imposed by the complex and comingled relationships between the salient characteristics of entrepreneurial action. This will reinforce future experiments' abilities to yield valid evidence about the causal effects of their manipulated variables – thus advancing understanding of entrepreneurship and its related phenomena.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jbusvent.2018.10.002>.

Acknowledgments

We extend our warmest thanks to Professor J. Robert Mitchell for his keen editorial guidance, to Jessica Chew for competent assistance in data collection, and to three anonymous reviewers for their excellent observations and suggestions on prior versions of this manuscript. All errors and omissions remain ours.

Appendix A

This appendix presents the methods we leveraged to identify 144 entrepreneurship research articles mobilizing experimental methods and to assess their mobilization of various validation strategies. Please consult Supplemental Online Appendix B for the full references to these articles.

A.1. Search procedures and inclusion/exclusion criteria

To identify a comprehensive corpus of entrepreneurship studies mobilizing experimental methods, we conducted a series of keyword searches mobilizing the web-archives of specific journals and publications databases. We then validated our search results in light of recently published inventories of entrepreneurship research experiments (specifically, Hsu et al., 2017a; and Kraus et al., 2016).

In light of prior reviews of this type and to center our results on articles that are most representative of entrepreneurship research's best practices of the era, we limited our online searches to journals listed in Thompson Reuter's Journal Citation Reports® (JCR) – Social Sciences Edition, 2016. Needless to say, we readily acknowledge that many insightful articles have appeared in journals not necessarily listed in the JCR. Yet building on the notion that JCR-indexing is an indicator of a journal's selectiveness and reputation (Nisonger, 2000; Stewart and Cotton, 2013), a number of prior reviews have used this sampling criterion as a strategy to identify published pieces that are most representative of a field's State-of-the-Art best practices (see Delgado García et al., 2015; Grégoire et al., 2011).

Using the search engines located on target journals and databases (explained below), we searched all published and in-press

articles that used the keyword “experimen*” anywhere in the text. We verified that the “*” allowed us to capture articles that used the related terms experiment, experiments, experimental, experimentally, and experimenting. As a further sampling strategy, we conducted a separate round of searches employing “conjoin*” to ensure that experiments termed as conjoint analyses would also be included. As readers may anticipate, these simple keyword searches yielded many articles that used different variants of the keywords but did not mobilize experimental methods. To sift these irrelevant pieces from our corpus, we obtained digital copies of the articles and used the search capabilities of Adobe®’s Acrobat Reader™ to systematically search for relevant instances of the term “experimen*” (as well as our own reading of the articles’ *Method* sections) to determine whether each met the following criteria:

1. Inclusion of at least one experimental manipulation/treatment designed to induce some sort of theoretically-relevant “difference” between groups of participants/stimuli/units of analyses.

This means that we parsed out a few studies that claimed to use an experiment but for which a careful reading showed that what the authors were effectively doing was using a hypothetical exercise that did *not* include a specific manipulation (e.g., Corbett, 2007). We also parsed out a few studies that claimed to use an experiment but for which a careful reading showed that what the authors were doing was reporting on an action-research intervention where their research design and analyses only included post hoc results – without any pre-test, control or comparison group (e.g. Friedman and Desivilya, 2010).

Conversely, however, and because they face the same validity challenges as other manipulation-based experiments, we kept in the corpus of analyzed articles the few studies that included manipulations that were not meant to test a focal hypothesis (whether explicitly formulated or not) but served instead to establish the generalizability of an observed effect across different conditions and/or experimental parameters. Many of these studies relied on games, lotteries or risky choices tasks that are common in behavioral economics (e.g., Arend, 2016; Chuah et al., 2016; Kuechle et al., 2016).

2. Research design that fosters either a controlled- or quasi-experiment

Because their inherent properties tend to pose radically different challenges than controlled and quasi-experiments (cf. Hsu et al., 2017a: p. 385; Shadish et al., 2002), we parsed out the natural experiments from our analyses. Natural experiments consist of studies where researchers do not deliberately manipulate a focal cause but take advantage of particular situations where one could observe “naturally occurring” variations in a specific factor of interest while other aspects of the situations remain similar and unchanging. By definition, these studies readily exhibit high levels of external and construct validity: conversely, their validity challenges more directly concern efforts to address issues of internal and statistical validity (e.g., selection biases that prevent them from ruling out alternate explanations). Accordingly, we excluded these studies from our samples.

That said, our sampling strategies include both controlled- and quasi-experiments. Needless to say, we readily acknowledge important differences between these two forms. By definition, quasi-experiments lack controlled experiments’ ability to randomly assigned participants to treatments or conditions. On the one hand, this threatens the internal validity of quasi-experiments, since the different treatments and control groups may suffer from confounding variables and selection biases and may not be immediately comparable prior to the manipulations’ onset. On the other hand, quasi-experiments offer experimenters the ability to select which groups will be assigned to different treatments (see Cook and Campbell, 1979; Shadish et al., 2002). In spite of these important differences, however, both controlled- and quasi-experiments mobilize active manipulation strategies and thus, face the same validity challenges of external and construct validity (Grant and Wall, 2009) – not to mention the tradeoffs discussed above with respect to mobilizing experimental methods in entrepreneurship research. To empirically examine whether entrepreneurship experiments and quasi-experiments mobilize different validation strategies, however, we report in Appendix D (below) results of a robustness check presenting the external and construct validation strategies of the quasi-experiments we identified in our samples.

3. Relevance with respect to entrepreneurship research.

To ensure the validity of our analyses, we built on prior research (Busenitz et al., 2003; Grégoire et al., 2011; Hsu et al., 2017a; Kraus et al., 2016) to parse out those studies which did not seek to contribute new academic knowledge on one or more topic(s) of entrepreneurship relevance, broadly conceived in light of the Academy of Management’s Entrepreneurship Division’s domain statement as of 2017 (<http://aom.org/Content.aspx?id=237#ent>). For instance, the application of this rule encouraged us to exclude Wu and Vosika (1983)’s experiment focusing on the effects of mail surveys’ length and sponsorship, but to keep Dennis Jr (2003)’s experiment on other survey properties; while both studies targeted similar research problematics (and samples of owner-operators of small and medium-size enterprises), only the latter made explicit arguments to the particular challenges of collecting survey responses among small-business owners.

A.2. Sampling frames

In order to examine whether validation practices might vary between experimental studies published in journals targeting different audiences, we conducted our searches and analyses across two sources of research articles: entrepreneurship-specific journals, and journals associated with neighboring sub-disciplines of applied psychology, behavioral economics, and the managerial and organizational sciences. This effectively yielded two samples of articles for our analyses, each with implicitly different levels of homogeneity (because of their different editorial specificity and target audiences).

A.2.1. Sample 1: experimental studies published in entrepreneurship-specific journals

For the first sample, we manually conducted a series of targeted keyword search on the websites of seven entrepreneurship journals listed in Thompson Reuter's Journal Citation Reports® (JCR) – Social Sciences Edition, 2016, and identified by [Stewart and Cotton \(2013\)](#) as the field's premier journals. The seven journals included were: 1) *Entrepreneurship and Regional Development* (ERD), 2) *Entrepreneurship Theory and Practice* (ETP), 3) the *International Small Business Journal* (ISBJ), 4) the *Journal of Business Venturing* (JBV), 5) the *Journal of Small Business Management* (JSBM), 6) *Small Business Economics* (SBE), and 7) the *Strategic Entrepreneurship Journal* (SEJ). Focusing on these seven journals ensures that we identified a corpus of articles that is highly representative of the field's “best practices” among the broader population of experimental studies specifically concerned with entrepreneurship-relevant phenomena.

While using the above search procedures on these seven journals yielded > 2500 hits, applying the inclusion/exclusion criteria described above brought this number down to 100 articles published or “in press” as of November 2017. To validate the representativeness of these sampling results, we compared them with the list of entrepreneurship experiment papers identified by [Kraus et al. \(2016\)](#) in their review: we found that with respect to the seven JCR-listed entrepreneurship journals identified above, our search procedures captured all the articles they had identified – and many more. We also compared our results with the list of entrepreneurship research experiments identified by [Hsu et al. \(2017a\)](#) in their recent review.⁵ Doing so revealed that for articles published up to 2015 (to account for the fact that their work was published online in the Fall of 2016 and might not include all papers for that year), our journal-specific keyword searches allowed us to capture 49 out of the 57 articles that they list (for these years and journals). Three articles were subsequently excluded because they were natural experiments, bringing the figures down to 46 out of 54. Of the eight we did not immediately identify, two did not have a focal manipulation, whereas the six others were simply not picked up by the databases' keyword searches (most likely because they did not expressly use the term “experiment” in the journals' search fields). We readily included these six studies (bringing the total to 52). More importantly, our procedures identified 28 more publications that were *not* identified by Hsu et al. (still for the same journals and years) – plus another 20 more relevant publications for the years 2016–17. All in all, these comparisons support the effectiveness of our search procedures – and suggest that our sample of 100 entrepreneurship-journal articles mobilizing controlled and quasi-experiments is not only representative, but also reasonably comprehensive. Please see Supplemental Online Appendix B, Section B.1. for the full references to the articles forming Sample 1.

A.2.2. Sample 2: experimental studies published in other social-science journals

To further assess our findings' robustness but also, to examine whether validation standards and practices might vary between entrepreneurship-specific journals and other journals from the broader social sciences, we built on [Hsu et al. \(2017a\)](#)'s review to include entrepreneurship research experiments published in journals targeting audiences in applied psychology, behavioral economics, and the managerial and organizational sciences.⁶ Applying the same inclusion/exclusion criteria, we started with a list of 24 articles from [Hsu et al. \(2017a\)](#).⁷ We added five otherwise pertinent articles listed by [Kraus et al. \(2016\)](#) in their review, but not included in [Hsu et al. \(2017a\)](#). Using the same systematic search procedures described above, we then conducted a series of searches in these journals to include experiments published or “in press” during the years 2016–17. Applying these procedures and the above inclusion/exclusion rules yielded 15 pertinent articles. All in all, these procedures generated a second sample of 44 entrepreneurship research articles mobilizing experiment and quasi-experiments and published in JCR-listed social-science journals not already included in the first sample – and targeting arguably broader audiences. Here again, we submit that these studies form a representative and comprehensive sample of entrepreneurship research experiments. Please see Supplemental Online Appendix B, Section B.2. for the references to Sample 2's articles.

Appendix B

Please consult Supplemental Online Appendix B for the full references to the 144 entrepreneurship research articles we analyzed as part of our structured literature review

⁵ Unfortunately, the text of their article does not disclose the specific manner how these authors identified the articles they reviewed: they only write that they “surveyed 29 academic journals” (p. 2), where the list of journals is adapted and expanded from a prior review of conjoint experiments by [Lohrke et al. \(2010\)](#), which derived their own list from [Fried \(2003\)](#). The tables in [Hsu et al. \(2017a\)](#) list 90 experimental studies reported in 87 different articles published in 19 journals.

⁶ These ten journals were the Academy of Management Journal (AMJ), Administrative Science Quarterly (ASQ), the American Economic Review (AER), the Journal of Applied Psychology (JAP), the Journal of Economic Behavior and Organization (JEB&O), the Journal of International Business Studies (JIBS), the Journal of Management (JMS), the Journal of Management Studies (JOMS), Management Science (MS), and the Strategic Management Journal (SMJ). Because entrepreneurship experiments had also appeared in Organization Science (OS) and Regional Studies (RS) in the interval (and these journals were included in Hsu et al. (2017)'s original list, we also included these journals to our sampling frame.

⁷ To make the comparison between our samples more consistent, we parsed out four articles they had identified but were published in journals not listed in the JCR (namely, the *Journal of Managerial Issues* and *Venture Capital*).

Appendix C

C.1. Establishing the reliability of our primary analyses

Consistent with State-of-the-Art recommendations for coding procedures (Duriau et al., 2007; Krippendorff, 2004; Neuendorf, 2002; Saldaña, 2015), we began by randomly selecting a subsample of 12 articles (eight from Sample 1 and four from Sample 2) which we used as our “training set”. Each author used the initial coding schemes to independently code all the articles in the set. We then compared the results, discussed the discrepancies and adjusted the coding schemes to better define our indicators. It is also at this stage that we identified illustrative examples of published entrepreneurship papers we could use as reference exemplars for our supplemental assessments. We then proceeded to re-code all articles in the training set in light of the revised coding scheme. Noticing that our respective results were converging, we agreed on a final code book, divided the articles from Sample 1 among ourselves, and began coding the articles in our respective set.

To establish the reliability of these procedures, we divided the coding of all articles in Sample 2 such that each article was independently coded by at least two people. Statistical analyses revealed that our coding procedures yielded valid and reliable results. For the validation strategies anchored on objective indicators directly expressed in the text of the analyzed articles, we obtained Cohen's K values that are 0.84 and above for external validity strategies and 0.81 and above for construct validity strategies (for an overall average of 0.89). We naturally resolved the discrepancies prior to reporting our results, yet these figures suggest that even if occasional mistakes might subsist, the results reported below offer meaningfully valid syntheses of the validation strategies mobilized in our two samples of articles.

C.2. Establishing the reliability of our supplemental analyses

To help ensure the consistency and reliability of our coding procedures for the two supplemental assessments, we developed a series of representative examples for each scale anchor. More specifically, we built on relevant exemplars of experimental studies – namely i) a study of opportunity identification based on a series of scenario manipulations (Grégoire and Shepherd, 2012), and ii) a typical conjoint experiment tapping into venture capitalists' (VCs') investment decision policies (Shepherd, 1999a, 1999b) – to illustrate different hypothetical articulations of these studies that would correspond to increases or decreases of their sample representativeness (external validity). In turn, we used these hypothetical examples as anchor points we systematically referred to in our rating of other entrepreneurship research experiments. Tables B.2a and B.2b present the details of these illustrative anchor points. To accommodate more refined nuances, we allowed for rating entries that fall in between scale anchors (e.g., 1.5, 2.5, 3.5 or 4.5).

We followed the same training and validation procedures described above to examine the reliability of our coding of the sample and procedural representativeness of the articles in Sample 2. Focusing on measures of inter-rater agreement (given the nature of the ratings), we obtained r_{wg} values of 0.86 and 0.93 – with comparative a_{wg} values of 0.71 a 0.87 for the two measures. These values are above the commonly-accepted threshold of 0.70 (see Brown and Hauenstein, 2005; LeBreton and Senter, 2008), indicating high levels of inter-rater agreement. Along the same line, calculations of average deviation (AD) indices (Smith-Crowe et al., 2013) yielded values of 0.20 and 0.18 for the two scales – figures that are well-below the *smaller-is-better* critical values advocated by these authors. Inter-rater correlations of 0.87 and 0.91 for the two scales, along with $ICC_{(1)}$ values of 0.87 and 0.89 and $ICC_{(K)}$ values of 0.93 and 0.94 indicate that the ratings were reliable across judges, and that the mean ratings (across raters) reliably distinguished different papers' levels of sample and procedural representativeness. Building on these results, we are confident of our procedures' reliability.

Table C.2a

Representative anchor points for assessing sample representativeness.

Scale anchors for sample representativeness	Hypothetical articulations of a scenario manipulation experiment to study opportunity identification (OI)	Hypothetical articulations of a conjoint experiment to study VCs' investment decisions
1 = very low (high-mismatch)	Doing OI research with undergraduate students that do not have much business and entrepreneurship experience, are not familiar with OI tasks, and are neither motivated nor specifically engaged in trying to identify opportunities because this is simply not what they focus on at this point in their life	Doing VC decision research on early-stage equity financing with undergraduate students who have neither experience, training nor particular motivations for doing what experienced VCs do
2 = low	Doing OI research with graduate MBAs that have some work experience, but are not necessarily knowledgeable, able, and/or motivated to engage in such tasks	Doing VC decision research on early-stage equity financing with graduate MBAs who have slightly more professional experience, but neither specific training nor particular motivation/are not doing this for a living
	Doing OI research with graduate MBAs that have received some specific training OI (thus business	Doing VC decision research on early-stage equity financing with finishing MBAs with a specialization

(continued on next page)

Table C.2a (continued)

Scale anchors for sample representativeness	Hypothetical articulations of a scenario manipulation experiment to study opportunity identification (OI)	Hypothetical articulations of a conjoint experiment to study VCs' investment decisions
3 = moderate: not great but defensible	experience + some familiarity and abilities with OI), but are not necessarily motivated to engage in such tasks Doing OI research with undergraduate students that have no prior business experience, but have received specific training (thus abilities and familiarity) and are actually engaged in some sort of entrepreneurship program where they have to engage in such tasks	in equity financing: the only missing thing is that they are not doing this for a living/this may not be relevant to them in their personal lives
4 = good but not perfect	Grégoire and Shepherd (2012)'s study Doing OI research with experienced entrepreneurs (familiar with the task and phenomena of opportunity identification) but that are <i>not</i> currently at the early stages of contemplating ideas for new entrepreneurial pursuits (even if they might vary in their openness to such pursuits).	Doing VC decision research on early-stage equity financing with novice VCs (thus perhaps less able/knowledgeable than experts) Doing VC decision research on early-stage equity financing with retired VCs (thus high experience and abilities, but no longer engaged in such activities)
5 = very high, near perfect	Doing OI research with experienced entrepreneurs that are currently at the early stages of contemplating new ideas for new entrepreneurial pursuits, for instance because they've enlisted in some sort of entrepreneurship program but have not yet finalized their ideation efforts (the only difference with an observational study would be that the experiment's task/material is different from what participants might do in real life)	Shepherd 1999ab's studies Doing VC decision research on early-stage equity financing with experienced VCs that are still engaged in such investment activities (the only difference with observational study is that the task/material is different from what people do in real life)

Table C.2b

Representative anchor points for assessing procedural representativeness.

Scale anchors	Hypothetical articulations of a scenario manipulation experiment to study opportunity identification (OI)	Hypothetical articulations of a conjoint experiment to study VCs' investment decisions
1 = very low (high-mismatch)	Doing the same study (Grégoire and Shepherd, 2012), but with opportunity scenarios that are not based on real cases AND focusing on outcomes that are tangential to the formation of OI beliefs (e.g., asking about anticipated market size or competitive superiority)	Very abstract conjoint with very abstract construct operationalizations Using a discrete conjoint task (presenting the profiles one at a time) when real-life phenomena imply an active comparison between different choices (and thus, where a choice-experiment would be more pertinent)
2 = low	Doing the same study (Grégoire and Shepherd, 2012), but with either: - research material that is now obsolete (e.g., studying 3D printing technologies now that these are ubiquitous) - focusing on outcomes that are tangential to the formation of OI beliefs (e.g., asking about anticipated market size or competitive superiority)	Doing a conjoint experiment using construct operationalizations that are somewhat more abstract and/or complex, derived primarily from prior theoretical efforts but that do not seem anchored on prior observational studies documenting the decision considerations used in real-life
3 = moderate: not great but defensible	Doing the same study (Grégoire and Shepherd, 2012), but with research material that is somewhat reduced, for instance using more abstract/less understandable descriptions (for instance), or basing the material on new technologies that have not yet been transferred/licensed	Shepherd 1999ab's studies Doing a conjoint experiment using variables that are explicitly based on prior observational studies (e.g., surveys) and make immediate sense to participants - variables that are easy to understand at face value); the main drawback is that conjoint profiles remain somewhat "dry", decontextualized and abstract (esp. (continued on next page)

Table C.2b (continued)

Scale anchors	Hypothetical articulations of a scenario manipulation experiment to study opportunity identification (OI)	Hypothetical articulations of a conjoint experiment to study VCs' investment decisions
4 = good but not perfect	Grégoire and Shepherd (2012)'s study Doing OI research with instructions, a task and research material that are all based on real-life (in this case, technology transfer); but in their case, the written material is a step back from experimenting the technologies in person	the articulation of the manipulations in simple terms of “high” and “low” Doing a conjoint-like experiment with richer, more articulated material, like a bus plan summary or short written story that does not simply provide highs and lows, but offer information that gives “flesh” to the manipulated variables
5 = very high, near perfect	Doing the same study “live” with real visits to R&D facilities and explanations describing the technologies, their origins, capabilities and market applications (richer material)	Doing a conjoint-like experiment with even richer, near complete material or more amenable material (like using a “live” set-up or manipulated videos)

Appendix D

This appendix reports the results of a series of robustness checks we conducted with sub-samples of experimental studies. These analyses help establish the general nature of the key findings and observations we report in the article.

D.1. External validation strategies

Table D.1 reports the mobilization of external validation strategies among two sub-samples of experiments that are more frequently conducted in entrepreneurship research. On the one hand, Panel A focuses on the decision experiments typically conducted through conjoint decisions, choice experiments and other policy capturing tasks.⁸ On the other hand, Panel B isolates the interpretation-focused experiments typically articulated through scenario manipulations of the information given to participants.⁹ By focusing on these two common forms of experiments (cf. Hsu et al. 2017; Kraus et al., 2016), we parcel out action-research and other types of interventions typically conducted in-situ with real-life samples of representative populations, and where the research manipulations, tasks and procedures tend to readily exhibit high levels of real-life representativeness. The sub-samples also parcel out the behavioral economic studies mobilizing more abstract and generic tasks (like games, lotteries, and other risky choices) that are claimed to exhibit high levels of psychological realism (they tap the same psychological processes) in spite of exhibiting otherwise low levels of external validity. Seen in this light, the two sub-samples constructed for Table D.1. allow us to focus on the more common “middle-of-the-road” experiments conducted in entrepreneurship research, net of other forms of experiments that might present different validity challenges. In turn, two main observations emerge from these analyses.

The results reported in Table D.1 Panel A indicate that a majority of decision experiments conducted in entrepreneurship research frequently mobilize pre-tests of their material (line 1). Many of these studies report having conducted pilot tests of their data collection procedures (line 2) And as we indicated in the main text, many of them report some sort of test-retest reliability assessment (line 6). There is also evidence that as a group, these studies tend to exhibit high levels of sample frame representativeness (line 8). Conversely, however, these studies' mobilization of other external validation strategies tends to be rare – and are either “on par” or slightly less frequent than what we observed for the entire samples. This is notably true for the use of post hoc checks for participant's attention, effort and engagement (line 3), for the conduct of multiple experiments or parallel studies (lines 4–5) and for the prevalence of specific comments about the magnitude of the observed relationships (line 8).

Results depicted in Table D.1 Panel B indicate that although many information manipulation experiments report having mobilized validation strategies like the conduct of pre-tests, pilot studies, post-experiment checks of attention, effort and/or engagement as well as the conduct of multiple studies and comments about the magnitude of observed relationships (particularly among studies

⁸ Such decision experiments typically mobilize simple stimuli (often amounting to one or a few words) for conveying the key information relevant to a particular decision scenario. Because of the scenarios' synthetic and attention-driving nature, participants can rapidly process the relevant information and perform the associated research task(s) (generally some sort of assessment of the scenario, expressing a preference or an intention, or making a choice). As such, decision experiments of this type can accommodate complex research designs that include concurrent variations and interactions across multiple (decision) parameters, yet the manipulations' synthetic form and the artificiality of the associated research task can sometimes raise concerns about the experiment's realism and ecological validity.

⁹ Information manipulation experiments typically rest on written scenarios but can also use graphic and audio/visual manipulations to vary the information presented to participants and document the subsequent effects of these variations. As such, information manipulations may bring together information contained in different cases, stories, scenarios and/or vignettes, graphic/visual depictions that are essentially static (like a drawing or a photography of a situation), or audio-only and audio/video recordings that showcase more dynamic movements and exchanges of information or even manipulations that engages other senses.

published in entrepreneurship journals), the numbers and frequencies for these strategies remain largely in line with.

Table D.1

The mobilization of external validation strategies in two subsets of experiments.

Panel A: Decision experiments (conjoint, choices, etc.), thus excluding interventions, lotteries, risky choices, and information manipulations Strategies	Number (%) – Sample 1 JCR®-listed entrepreneurship journals				Number (%)– Sample 2 JCR®-listed social-science journals			
	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)
	n = 47	n = 18	n = 17	n = 12	n = 12	n = 5	n = 4	n = 3
1. Pre-test of an experiment's research material to establish that it is representative	29 (61.70)	12 (66.67)	11 (64.71)	6 (50.00)	9 (75.00)	4 (80.00)	3 (75.00)	2 (66.67)
2. Pilot-test of an experiment's entire data collection procedures	16 (34.04)	6 (33.33)	4 (23.53)	6 (50.00)	6 (50.00)	3 (60.00)	2 (50.00)	1 (33.33)
3. Post-experiment's checks to assess participants' attention, effort and/or engagement	3 (6.38)	0	1 (5.88)	2 (16.67)	1 (8.33)	0	1 (25.00)	0
4. Multiple experiments within single paper	3 (6.38)	0	1 (5.88)	2 (16.67)	0	0	0	0
5. Parallel study using other data collection methods	5 (10.64)	1 (5.56)	3 (17.65)	1 (8.33)	1 (8.33)	1 (20.00)	0	0
6. Test-retest reliability assessment	28 (59.57)	11 (61.11)	9 (52.94)	8 (66.66)	11 (91.67)	5 (100.0)	4 (100.0)	2 (66.67)
7. Reporting of effect sizes and/or comments about the magnitude of observed relationships	12 (25.53)	5 (27.78)	3 (17.65)	4 (33.33)	1 (8.33)	1 (20.00)	0	0
8. Sample frame representativeness (average)	4.39	4.58	4.24	4.33	4.52	4.70	4.44	4.33
9. Procedural representativeness (average)	2.65	2.75	2.32	2.95	2.83	2.70	3.13	2.67
Panel B:	Number (%) – Sample 1				Number (%)– Sample 2			
Interpretative experiments based on information manipulations, thus excluding interventions, lotteries, risky choices, and decision experiments (choices, conjoint, etc.) Strategies	JCR®-listed entrepreneurship journals				JCR®-listed social-science journals			
	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)
	n = 34	n = 14	n = 9	n = 11	n = 20	n = 4	n = 3	n = 13
1. Pre-test of an experiment's research material to establish that it is representative	13 (38.24)	8 (57.14)	1 (11.11)	4 (36.36)	4 (20.00)	1 (25.00)	1 (33.33)	2 (15.38)
2. Pilot-test of an experiment's entire data collection procedures	6 (17.65)	1 (7.14)	2 (22.22)	3 (27.27)	1 (5.00)	0	0	1 (7.69)
3. Post-experiment's checks to assess participants' attention, effort and/or engagement	4 (11.76)	2 (14.29)	0	2 (18.18)	1 (5.00)	0	0	1 (7.69)
4. Multiple experiments within single paper	12 (35.29)	4 (28.57)	4 (44.44)	4 (36.36)	5 (25.00)	0	2 (66.67)	3 (23.08)
5. Parallel study using other data collection methods	3 (8.82)	1 (7.14)	1 (11.11)	1 (9.09)	9 (45.0)	3 (75.00)	1 (33.33)	5 (38.46)
6. Test-retest reliability assessment	0	0	0	0	1 (5.00)	0	0	1 (7.69)
7. Reporting of effect sizes and/or comments about the magnitude of observed relationships	16 (47.06)	4 (28.57)	6 (66.67)	6 (54.55)	10 (50.00)	2 (50.00)	1 (33.33)	7 (53.85)
	3.19	3.71	2.22	3.32	2.91	2.38	3.04	3.04

(continued on next page)

Table D.1 (continued)

Panel A: Decision experiments (conjoint, choices, etc.), thus excluding interventions, lotteries, risky choices, and information manipulations Strategies	Number (%) – Sample 1 JCR®-listed entrepreneurship journals				Number (%)– Sample 2 JCR®-listed social-science journals			
	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)
	n = 47	n = 18	n = 17	n = 12	n = 12	n = 5	n = 4	n = 3
8. Sample frame representativeness (average)								
9. Procedural representativeness (average)	3.34	3.64	2.89	3.32	3.16	2.88	2.42	3.43

figures we reported in the main text – and which included a more diver set of experimental manipulations.

D.2. Construct validation strategies

Panels C and D in Table D.2 below break down the construct validation results for the same two subsets decision experiments and information manipulations. By and large, these analyses reveal that construct validation strategies are infrequently mobilized among decision experiments (see Panel C). Indeed, many decision experiments do not conduct any such tests. Because decision experiments count for a sizable portion of the total number of entrepreneurship research experiments published in both entrepreneurship journals (47 out of 100 studies) and social-science journals (20 out of 44), the less frequent mobilization of these strategies explain the overall observations reported in the main text of our article. Along this line, evidence displayed in Panel D indicates that if strategies to examine construct validity issues are more prevalent among interpretive experiments anchored on scenario manipulations (with figures that tend to be larger than those reported in the main article), the mobilization of such strategies remains far from systematic. Here again, a number of such experiments are published without mobilizing any strategy to establish the validity of their manipulations.

D.3. External and construct validation strategies in quasi-experiments

As a further robustness check to examine potential differences among the validation strategies mobilized in different subsets of entrepreneurship research experiments, Table D.3 below reports the validation strategies mobilized among the quasi-experiments in our two samples (i.e., 22 quasi-experiments from Sample 1 and six from Sample 2). Comparing the pattern of validation strategies mobilized in quasi-experiments with those observed for the two main samples (cf. Tables 2 and 3 above), we observe a few salient differences. For instance, quasi-experiments mobilize pre-test to establish representativeness less frequently than what was observed in the overall samples (17.86% vs. 43% and 31.82 for Samples 1 and 2, respectively). Likewise, smaller proportions of quasi-experiments mobilize pilot-tests (10.71% vs. 23% and 18.18%) or discuss the magnitude of observed effects (25% vs. 31% and 38.64%). Yet a higher proportion of articles mobilize multiple experiments within a single article (21.43% vs. 15% and 11.36%). These differences might be due to the fact that quasi-experiments tend to mobilize research procedures and circumstances that readily exhibit higher levels of external validity, such that researchers might not feel the need to report more specific validity tests. Indeed, 10 of the 28 quasi-experiments we identified (36%) consist of studies using an entrepreneurship course to test the effects of a particular form of training on students' intentions. This constitutes an ecologically valid and realistic approach to evaluate this type of training. Along this line, we noted that the average ratings of procedural representativeness for the quasi-experiments (3.81) was sensibly higher than that obtained for Sample 1 (3.11) and Sample 2 (3.26). That being said, quasi-experiment's mobilization of construct validation strategies is on par with that observed in the overall samples, with 57.15% of quasi-experiments not reporting any relevant tests.

Table D.2

The mobilization of construct validation strategies in entrepreneurship experiments.

Construct validation strategies	Number (%) – Sample 1 JCR®-listed entrepreneurship journals				Number (%)– Sample 2 JCR®-listed social-science journals			
	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)
	n = 47	n = 18	n = 17	n = 12	n = 12	n = 5	n = 4	n = 3
Panel C: Decision experiments		0	2 (11.76)	2 (16.67)		0	1 (25.00)	0

(continued on next page)

Table D.2 (continued)

	Construct validation strategies	Number (%) – Sample 1 JCR®-listed entrepreneurship journals				Number (%)– Sample 2 JCR®-listed social-science journals			
		Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)
(conjoint, choice, etc.)	1. Separate construct validity study	4 (8.51)				1 (8.33)			
	2. Post-manipulation checks of material's understanding	2 (4.26)	0	2 (11.76)	0	0	0	0	0
	3. Post-experiment checks of manipul. effectiveness	12 (25.53)	3 (16.67)	4 (23.53)	5 (41.67)	3 (25.00)	1 (20.00)	1 (25.00)	1 (33.33)
	4. No construct validation strategy	31 (65.96)	15 (83.33)	10 (58.82)	6 (50.00)	9 (75.00)	4 (80.00)	3 (75.00)	2 (66.67)
Panel D:		n = 34	n = 14	n = 9	n = 11	n = 20	n = 4	n = 3	n = 13
Interpretative experiments based on information manipulations	1. Separate construct validity study	9 (26.47)	2 (14.29)	4 (44.44)	3 (27.27)	6 (30.00)	1 (25.00)	1 (33.33)	4 (30.77)
	2. Post-manipulation checks of material's understanding	5 (14.71)	0	3 (33.33)	2 (18.18)	5 (25.00)	2 (50.00)	0	3 (23.08)
	3. Post-experiment checks of manipul. effectiveness	22 (64.70)	11 (78.57)	4 (44.44)	7 (63.64)	8 (40.00)	2 (50.00)	0	6 (46.15)
	4. No construct validation strategy	6 (17.65)	2 (14.29)	1 (11.11)	3 (27.27)	7 (35.00)	0	2 (66.67)	5 (38.46)

Table D.3

Robustness check: the mobilization of external and construct validation strategies in entrepreneurship research quasi-experiments.

Panel E: Quasi-experiments only		Number (%) or average ratings (standard deviation) – Sample 1 and Sample 2 combined			
Sub-dimensions	Strategies	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)
		n = 28	n = 11	n = 3	n = 14
Ecological validity/ mundane realism	1. Pre-test of an experiment's research material to establish that it is representative	5 (17.86)	4 (36.36)	0	1 (7.14)

(continued on next page)

Table D.3 (continued)

Panel E: Quasi-experiments only		Number (%) or average ratings (standard deviation) – Sample 1 and Sample 2 combined			
Sub-dimensions	Strategies	Overall	Prior to 2010	2010–14 (incl.)	2015–17 (incl.)
		n = 28	n = 11	n = 3	n = 14
Experimental realism	2. Pilot-test of an experiment's entire data collection procedures	3 (10.71)	1 (9.09)	0	2 (14.29)
	3. Post-experiment's checks to assess participants' attention, effort and/or engagement	1 (4.55)	1 (9.09)	0	0
Robustness	4. Multiple experiments within single paper	6 (21.43)	3 (27.27)	0	3 (21.43)
	5. Parallel study using other data collection methods	0	0	0	0
Relevance	6. Use of manipulation(s) to assess the stability of another relationship across different conditions	0	0	0	0
	7. Test-retest reliability assessment	0	0	0	0
	8. Reporting of effect sizes and/or comments about the magnitude of observed relationships in terms of their underlying variables and metrics	7 (25.00)	2 (18.18)	1 (33.33)	4 (28.57)
Overall assessments of external validity	9. Sample frame representativeness	3.88 (1.31)	4.23 (1.25)	4.33 (1.25)	3.52 (1.44)
	10. Procedural representativeness	3.81 (1.47)	3.95 (1.40)	4.67 (0.58)	3.53 (1.63)
Construct validity	11. Separate construct validity study	1 (3.57)	0	0	1 (7.14)
	12. Post-manipulation checks of material's understanding	2 (7.14)	1 (9.09)	0	1 (7.14)
	13. Post-experiment checks of manipulation awareness	0	0	0	0
	14. Post-experiment checks of manipulation effectiveness	11 (39.29)	8 (72.72)	0	3 (21.43)
	15. No construct validation strategy	16 (57.15)	3 (27.27)	3 (100)	10 (71.43)

Appendix E

Please consult Supplemental Online Appendix E for an extensive step-by-step guide of design and assessment strategies relevant for conducting entrepreneurship research experiments.

References*

- Aguinis, H., Lawal, S.O., 2012. Conducting field experiments using eLancing's natural environment. *J. Bus. Ventur.* 27 (4), 493–505.
- Aguinis, H., Ramani, R.S., Alabduljader, N., 2018. What you see is what you get? Enhancing methodological transparency in management research. *Acad. Manag. Ann.* 12 (1), 83–110.
- Alferes, V.R., 2012. *Methods of Randomization in Experimental Design*. Sage Publications, Thousand Oaks, CA.
- Arend, R.J., 2016. Entrepreneurs as sophisticated iconoclasts: rational rule-breaking in an experimental game. *J. Small Bus. Manag.* 54 (1), 319–340.
- Aronson, E., Carlsmith, J.M., 1968. Experimentation in social psychology. In: Lindzey, G., Aronson, E. (Eds.), *The Handbook of Social Psychology*. Vol. 2. Addison-Wesley, Reading, MA, pp. 1–79.
- Aronson, E., Wilson, T.D., Akert, R.M., 1994. *Social Psychology: The Heart and the Mind*. Harper-Collins, New York, NY.
- Baron, R.A., 1998. Cognitive mechanisms in entrepreneurship: why and when entrepreneurs think differently than other people. *J. Bus. Ventur.* 13 (4), 275–294.
- Bello, D., Leung, K., Radebaugh, L., Tung, R.L., Van Witteloostuijn, A., 2009. From the editors: student samples in international business research. *J. Int. Bus. Stud.* 40 (3), 361–364.
- Bergh, D.D., Sharp, B.M., Aguinis, H., Li, M., 2017. Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strateg. Organ.* 15 (3), 423–436.

* Note: For the complete list of entrepreneurship research experiment articles that we analyzed in our review, please consult Supplemental Online Appendix B.

- Berinsky, A.J., Huber, G.A., Lenz, G.S., 2012. Evaluating online labor markets for experimental research: Amazon. Com's Mechanical Turk. *Polit. Anal.* 20 (3), 351–368.
- Bettis, R.A., Ethiraj, S., Gambardella, A., Helfat, C., Mitchell, W., 2016. Creating repeatable cumulative knowledge in strategic management. *Strateg. Manag. J.* 37 (2), 257–261.
- Brewer, M.B., 2000. Research design and issues of validity. In: Reis, H.T., Judd, C.M. (Eds.), *Handbook of Research Methods in Social and Personality Psychology*. Cambridge University Press, Cambridge, UK, pp. 3–16.
- Brown, R.D., Hauenstein, N.M., 2005. Interrater agreement reconsidered: an alternative to the rwg indices. *Organ. Res. Methods* 8 (2), 165–184.
- Busenitz, L.W., West III, G.P., Shepherd, D., Nelson, T., Chandler, G.N., Zacharakis, A., 2003. Entrepreneurship research in emergence: past trends and future directions. *J. Manag.* 29 (3), 285–308.
- Camerer, C.F., Fehr, E., 2006. When does “economic man” dominate social behavior? *Science* 311 (5757), 47–52.
- Cardon, M.S., Zietsma, C., Saparito, P., Matherne, B.P., Davis, C., 2005. A tale of passion: new insights into entrepreneurship from a parenthood metaphor. *J. Bus. Ventur.* 20 (1), 23–45.
- Chen, X.P., Yao, X., Kotha, S., 2009. Entrepreneur passion and preparedness in business plan presentations: a persuasion analysis of venture capitalists' funding decisions. *Acad. Manag. J.* 52 (1), 199–214.
- Cheung, J.H., Burns, D.K., Sinclair, R.R., Sliter, M., 2017. Amazon Mechanical Turk in organizational psychology: an evaluation and practical recommendations. *J. Bus. Psychol.* 32 (4), 347–361.
- Chuah, S.H., Hoffmann, R., Ramasamy, B., Tan, J.H., 2016. Is there a spirit of overseas Chinese capitalism? *Small Bus. Econ.* 47 (4), 1095–1118.
- Clarke, J.S., Cornelissen, J.P., Healey, M., 2018. Actions speak louder than words: how figurative language and gesturing in entrepreneurial pitches influences investment judgments. *Acad. Manag. J.* <https://doi.org/10.5465/amj.2016.1008>. (in press).
- Clifford, S., Jewell, R.M., Waggoner, P.D., 2015. Are samples drawn from Mechanical Turk valid for research on political ideology? *Res. Politics* 2 (4), 1–15.
- Colquitt, J.A., 2008. Publishing laboratory research in AMJ: a question of when, not if. *Acad. Manag. J.* 51 (4), 610–620.
- Cook, T.D., Campbell, D.T., 1979. *Quasi-experimentation*. Houghton Mifflin, Boston, MA.
- Corbett, A.C., 2007. Learning asymmetries and the discovery of entrepreneurial opportunities. *J. Bus. Ventur.* 22 (1), 97–118.
- Crano, W.D., Brewer, M.A., 2002. *Principles and Methods of Social Research*, 2nd ed. Lawrence Erlbaum, Mahwah, NJ.
- Crook, T.R., Shook, C.L., Morris, M.L., Madden, T.M., 2010. Are we there yet? An assessment of research design and construct measurement practices in entrepreneurship research. *Organ. Res. Methods* 13 (1), 192–206.
- Cuervo-Cazurra, A., Andersson, U., Brannen, M.Y., Nielsen, B.B., Reuber, A.R., 2016. From the editors: can I trust your findings? Ruling out alternative explanations in international business research. *J. Int. Bus. Stud.* 47 (8), 881–897.
- Davidsson, P., 2006. Method challenges and opportunities in the psychological study of entrepreneurship. In: Baum, J.R., Frese, M., Baron, R.A. (Eds.), *The Psychology of Entrepreneurship*. Erlbaum, Mahwah, NJ, pp. 287–323.
- Davidsson, P., 2016. *Researching Entrepreneurship: Conceptualization and Design*. Springer.
- Delgado García, J.B., Quevedo Puente, E., Blanco Mazagatos, V., 2015. How affect relates to entrepreneurship: a systematic review of the literature and research agenda. *Int. J. Manag. Rev.* 17 (2), 191–211.
- Dennis Jr., W.J., 2003. Raising response rates in mail surveys of small business owners: results of an experiment. *J. Small Bus. Manag.* 41 (3), 278–295.
- DeTienne, D.R., Shepherd, D.A., De Castro, J.O., 2008. The fallacy of “only the strong survive”: the effects of extrinsic motivation on the persistence decisions for under-performing firms. *J. Bus. Ventur.* 23 (5), 528–546.
- DeVellis, R.F., 2016. *Scale Development: Theory and Applications*. Vol. 26 Sage Publications, Thousand Oaks, CA.
- Dreyfuss, E., 2018. A bot panic hits Amazon's Mechanical Turk. *Wired magazine*. <https://www.wired.com/story/amazon-mechanical-turk-bot-panic/>.
- Drover, W.W., Wood, M.S., Fassin, Y., 2014. Take the money or run? Investors' ethical reputation and entrepreneurs' willingness to partner. *J. Bus. Ventur.* 29 (6), 723–740.
- Duriau, V.J., Rege, R.K., Pfarrer, M.D., 2007. A content analysis of the content analysis literature in organization studies: research themes, data sources, and methodological refinements. *Organ. Res. Methods* 10 (1), 5–34.
- Dvir, T., Eden, D., Avolio, B.J., Shamir, B., 2002. Impact of transformational leadership on follower development and performance: a field experiment. *Acad. Manag. J.* 45 (4), 735–744.
- Edwards, J.R., Berry, J.W., 2010. The presence of something or the absence of nothing: increasing theoretical precision in management research. *Organ. Res. Methods* 13 (4), 668–689.
- Fisher, R.J., 1993. Social desirability bias and the validity of indirect questioning. *J. Consum. Res.* 20 (2), 303–315.
- Fitzsimmons, J.R., Douglas, E.J., 2011. Interaction between feasibility and desirability in the formation of entrepreneurial intentions. *J. Bus. Ventur.* 26 (4), 431–440.
- Fried, V., 2003. Defining a forum for entrepreneurship scholars. *J. Bus. Ventur.* 18 (1), 1–11.
- Friedman, V.J., Desivilya, H., 2010. Integrating social entrepreneurship and conflict engagement for regional development in divided societies. *Entrep. Reg. Dev.* 22 (6), 495–514.
- Furr, M., 2011. *Scale Construction and Psychometrics for Social and Personality Psychology*. Sage Publications, Thousand Oaks, CA.
- Gartner, W.B., 2001. Is there an elephant in entrepreneurship? Blind assumptions in theory development. *Enterp. Theory Pract.* 25 (4), 27–39.
- Goldfarb, B., King, A.A., 2016. Scientific apophenia in strategic management research: significance tests & mistaken inference. *Strateg. Manag. J.* 37 (1), 167–176.
- Gordon, M.E., Slade, L.A., Schmitt, N., 1987. Student guinea pigs: porcine predictors and particularistic phenomena. *Acad. Manag. Rev.* 12 (1), 160–163.
- Grant, A.M., Wall, T.D., 2009. The neglected science and art of quasi-experimentation: why-to, when-to, and how-to advice for organizational researchers. *Organ. Res. Methods* 12 (4), 653–686.
- Grégoire, D.A., Shepherd, D.A., 2012. Technology-market combinations and the identification of entrepreneurial opportunities: an investigation of the opportunity-individual nexus. *Acad. Manag. J.* 55 (4), 753–785.
- Grégoire, D.A., Shepherd, D.A., Schurer Lambert, L., 2010. Measuring opportunity-recognition beliefs: Illustrating and validating an experimental approach. *Organ. Res. Methods* 13 (1), 114–145.
- Grégoire, D.A., Corbett, A.C., McMullen, J.S., 2011. The cognitive perspective in entrepreneurship: an agenda for future research. *J. Manag. Stud.* 48 (6), 1443–1477.
- Gupta, V.K., Turban, D.B., Bhawe, N.M., 2008. The effect of gender stereotype activation on entrepreneurial intentions. *J. Appl. Psychol.* 93 (5), 1053–1061.
- Gustafsson, V., 2006. *Entrepreneurial Decision-making: Individuals, Tasks and Cognitions*. Edward Elgar Publishing.
- Highhouse, S., 2009. Designing experiments that generalize. *Organ. Res. Methods* 12 (3), 554–566.
- Hinkin, T.R., 1998. A brief tutorial on the development of measures for use in survey questionnaires. *Organ. Res. Methods* 1 (1), 104–121.
- Hinkin, T.R., Tracey, J.B., 1999. An analysis of variance approach to content validation. *Organ. Res. Methods* 2 (2), 175–186.
- Holm, H.J., Oppen, S., Nee, V., 2013. Entrepreneurs under uncertainty: An economic experiment in China. *Manag. Sci.* 59 (7), 1671–1687.
- Hsu, D.K., Simmons, S.A., Wieland, A.M., 2017a. Designing entrepreneurship experiments: a review, typology, and research agenda. *Organ. Res. Methods* 20 (3), 379–412.
- Hsu, D.K., Wiklund, J., Cotton, R.D., 2017b. Success, failure, and entrepreneurial reentry: an experimental assessment of the veracity of self-efficacy and prospect theory. *Enterp. Theory Pract.* 41 (1), 19–47.
- Iselin, E., 1989. The impact of information diversity on information overload effects in unstructured managerial decision making. *J. Inf. Sci.* 15 (3), 163–173.
- Jones, O., Gatrell, C., 2014. Editorial: the future of writing and reviewing for IJMR. *Int. J. Manag. Rev.* 16 (3), 249–264.
- Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47 (2), 263–291.
- Katz, J., Gartner, W.B., 1988. Properties of emerging organizations. *Acad. Manag. Rev.* 13 (3), 429–441.
- Keith, M.G., Harms, P.D., 2016. Is Mechanical Turk the answer to our sampling woes? *Ind. Organ. Psychol.* 9 (1), 162–167.
- Kerr, W.R., Lerner, J., Schoar, A., 2011. The consequences of entrepreneurial finance: evidence from angel financings. *Rev. Financ. Stud.* 27 (1), 20–55.
- Kirk, R.E., 2013. *Experimental Design*, 4th edition. Sage Publications, Thousand Oaks, CA.

- Koudstaal, M., Sloof, R., van Praag, C.M., 2015. Risk, uncertainty and entrepreneurship: evidence from a large lab-in-the-field experiment. *Manag. Sci.* 62 (10), 2897–2915.
- Kraus, S., Meier, F., Niemand, T., 2016. Experimental methods in entrepreneurship research: the status quo. *Int. J. Entrep. Behav. Res.* 22 (6), 958–983.
- Krippendorff, K., 2004. Reliability in content analysis. *Hum. Commun. Res.* 30 (3), 411–433.
- Kuechle, G., Boulu-Reshef, B., Carr, S.D., 2016. Prediction and control-based strategies in entrepreneurship: the role of information. *Strateg. Entrep. J.* 10 (1), 43–64.
- Lautenschlager, G.J., Flaherty, V.L., 1990. Computer administration of questions: More desirable or more social desirability? *J. Appl. Psychol.* 75 (3), 310–314.
- LeBreton, J.M., Senter, J.L., 2008. Answers to 20 questions about interrater reliability and interrater agreement. *Organ. Res. Methods* 11 (4), 815–852.
- Lohrke, F.T., Holloway, B.B., Woolley, T.W., 2010. Conjoint analysis in entrepreneurship research: a review and research agenda. *Organ. Res. Methods* 13 (1), 16–30.
- Maxwell, S.E., Delaney, H.D., 2004. *Designing Experiments and Analyzing Data*, Second edition. Psychology Press Taylor & Francis Group, New York, NY.
- McClelland, G.H., 2000. Nasty data: Unruly, ill-mannered observations can ruin your analysis. In: Reis, H.T., Judd, C.M. (Eds.), *Handbook of Research Methods in Social and Personality Psychology*. Cambridge University Press, Cambridge, UK, pp. 393–411.
- McGee, J.E., Peterson, M., Mueller, S.L., Sequeira, J.M., 2009. Entrepreneurial self-efficacy: refining the measure. *Enterp. Theory Pract.* 33 (4), 965–988.
- McKelvie, A., Haynie, J.M., Gustavsson, V., 2011. Unpacking the uncertainty construct: implications for entrepreneurial action. *J. Bus. Ventur.* 26 (3), 273–292.
- McMullen, J.S., Dimov, D., 2013. Time and the entrepreneurial journey: the problems and promise of studying entrepreneurship as a process. *J. Manag. Stud.* 50 (8), 1481–1512.
- McMullen, J.S., Kier, A.S., 2016. Trapped by the entrepreneurial mindset: opportunity seeking and escalation of commitment in the Mount Everest disaster. *J. Bus. Ventur.* 31 (6), 663–686.
- Mead, R., Gilmour, S.G., Mead, A., 2012. *Statistical Principles for the Design of Experiments: Applications to Real Experiments*. Cambridge University Press, Cambridge, UK.
- Merriam-Webster's Online Dictionary, 2017. <https://www.merriam-webster.com/dictionary/valid>.
- Mueller, J., Melwani, S., Loewenstein, J., Deal, J.J., 2018. Reframing the decision-makers' dilemma: towards a social context model of creative idea recognition. *Acad. Manag. J.* 61 (1), 94–110.
- Neuendorf, K.A., 2002. *The Content Analysis Guidebook*. Sage, Thousand Oaks, CA.
- Nisonger, T.E., 2000. Use of the Journal Citation Reports for serials management in research libraries: an investigation of the effect of self-citation on journal rankings in library and information science and genetics. *Coll. Res. Libr.* 61 (3), 263–275.
- Olejnik, S., Algina, J., 2003. Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychol. Methods* 8 (4), 434–447.
- Perdue, B.C., Summers, J.O., 1986. Checking the success of manipulations in marketing experiments. *J. Mark. Res.* 23 (4), 317–326.
- Pierce, C.A., Block, R.A., Aguinis, H., 2004. Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educ. Psychol. Meas.* 64 (6), 916–924.
- Saldaña, J., 2015. *The Coding Manual for Qualitative Researchers*. Sage, Thousand Oaks, CA.
- Sánchez, J.C., 2013. The impact of an entrepreneurship education program on entrepreneurial competencies and intention. *J. Small Bus. Manag.* 51 (3), 447–465.
- Shadish, W.R., Cook, T.D., Campbell, D.T., 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin Company, Boston, MA.
- Shane, S., Venkataraman, S., 2000. The promise of entrepreneurship as a field of research. *Acad. Manag. Rev.* 25 (1), 217–226.
- Shapiro, D.N., Chandler, J., Mueller, P.A., 2013. Using Mechanical Turk to study clinical populations. *Clin. Psychol. Sci.* 1 (2), 213–220.
- Shepherd, D.A., 1999a. Venture capitalists' assessment of new venture survival. *Manag. Sci.* 45 (5), 621–632.
- Shepherd, D.A., 1999b. Venture capitalists' introspection: a comparison of "in use" and "espoused" decision policies. *J. Small Bus. Manag.* 37 (2), 76–87.
- Smith, E.R., 2000. Research design. In: Reis, H.T., Judd, C.M. (Eds.), *Handbook of Research Methods in Social and Personality Psychology*. Cambridge University Press, Cambridge, UK, pp. 17–39.
- Smith-Crowe, K., Burke, M.J., Kouchaki, M., Signal, S.M., 2013. Assessing interrater agreement via the average deviation index given a variety of theoretical and methodological problems. *Organ. Res. Methods* 16 (1), 127–151.
- Souitaris, V., Zerbinati, S., Al-Laham, A., 2007. Do entrepreneurship programmes raise entrepreneurial intention of science and engineering students? The effect of learning, inspiration and resources. *J. Bus. Ventur.* 22 (4), 566–591.
- Stevens, C.K., 2011. Questions to consider when selecting student samples. *J. Supply Chain Manag.* 47 (3), 19–21.
- Stevenson, R.M., Ciuchta, M.P., Letwin, C., Dinger, J.M., Vancouver, J.B., 2018. Out of control or right on the money? Funder self-efficacy and crowd bias in equity crowdfunding. *J. Bus. Ventur.* (in press).
- Stewart, A., Cotton, J., 2013. Making sense of entrepreneurship journals: journal rankings and policy choices. *Int. J. Entrep. Behav. Res.* 19 (3), 303–323.
- Sutcliffe, K.M., Weick, K.E., 2008. Information overload revisited. In: Hodgkinson, G.P., Starbuck, W.H. (Eds.), *The Oxford Handbook of Organizational Decision Making*. Oxford University Press, Oxford, UK, pp. 56–75.
- Townsend, D., Hunt, R., McMullen, J.S., Sarasvathy, S., 2018. Uncertainty, knowledge problems, and entrepreneurial action. *Acad. Manag. Ann.* <https://doi.org/10.5465/annals.2016.0109>.
- Tranfield, D., Denyer, D., Smart, P., 2003. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *Br. J. Manag.* 14 (3), 207–222.
- Urbig, D., 2016. Student samples in entrepreneurship research. <http://blog.diemo.de/index.php/archives/3-Student-samples-in-entrepreneurship-research.html>.
- Volery, T., Müller, S., Oser, F., Naepflin, C., Rey, N., 2013. The impact of entrepreneurship education on human capital at upper-secondary level. *J. Small Bus. Manag.* 51 (3), 429–446.
- Webster's New Dictionary of Synonyms, 1984. Merriam-Webster, Springfield, MA.
- Weinberger, E., Wach, D., Stephan, U., Wegge, J., 2018. Having a creative day: understanding entrepreneurs' daily idea generation through a recovery lens. *J. Bus. Ventur.* 33 (1), 1–19.
- West, S.G., Biesanz, J.C., Pitts, S.C., 2000. Causal inference and generalization in field settings: Experimental and quasi-experimental designs. In: Reis, H.T., Judd, C.M. (Eds.), *Handbook of Research Methods in Social and Personality Psychology*. Cambridge University Press, Cambridge, UK, pp. 40–84.
- Wilson, T.D., Aronson, E., Carlsmith, K., 2010. The art of laboratory experimentation. In: Fiske, S.T., Gilbert, D.T., Lindzey, G. (Eds.), *Handbook of Social Psychology*. Vol. 1. pp. 51–81 Part 1, Ch. 2.
- Woo, S.E., Keith, M., Thornton, M.A., 2015. Amazon mechanical Turk for industrial and organizational psychology: advantages, challenges, and practical recommendations. *Ind. Organ. Psychol.* 8 (2), 171–179.
- Wood, M.S., Williams, D.W., Drover, W., 2017. Past as prologue: Entrepreneurial inaction decisions and subsequent action judgments. *J. Bus. Ventur.* 32 (1), 107–127.
- Wu, B.T., Vosika, J., 1983. Improving primary research: an experimental study of mail survey response. *J. Small Bus. Manag.* 21 (2), 30.