

Moving beyond the baseline: Exploring the potential of experi- ments in language research

Shea X. Fan
Anne-Wil Harzing

Version August 2019

Accepted for Horn, S., Lecomte, P. and Tietze, S. (eds) 2020 *Understanding Multilingual Workplaces: Methodological, Empirical and Pedagogic Perspectives*, Routledge.

Copyright © 2019 Shea X. Fan, Anne-Wil Harzing
All rights reserved.

Prof. Anne-Wil Harzing
Middlesex University
The Burroughs, Hendon
London NW4 4BT
Email: anne@harzing.com
Web: www.harzing.com

**Moving beyond the baseline:
Exploring the potential of experiments in language research**

Shea Fan, RMIT University

Anne-Wil Harzing, Middlesex University

Introduction

In the last 30 years language research in International Management has changed from being a niche topic to occupying the centre stage. Researchers are thus starting to look back at what has been achieved and to consider how to move this stream of research forward (see e.g. Karhunen, Kankaanranta, Louhiala-Salminen, & Piekkari, 2018; Tenzer, Terjesen, & Harzing, 2017). This chapter also contributes to this goal, but with a narrower focus, i.e. analysing the use of experimental designs in language research in International Management. In the rest of this chapter, we will use the term language research to mean language research in International Management.

One of the key purposes of research is developing a theory and using it to make predictions. Theory is “about the connections among phenomena, a story about why acts, events, structure, and thoughts occur” (Sutton & Staw, 1995, p. 378). To test whether the theory we have developed is plausible, we need evidence of causality, specifically 1) covariation between cause and effect, (2) temporal precedence of the cause, and (3) elimination of alternative explanations for the causal relationship (Cook & Campbell, 1979). Compared with surveys and case studies, an obvious advantage of experimental design is its ability to establish causality through a process called randomization. By randomly assigning participants to treatment and control groups, an experimental design can ensure that no unmeasured variables are systemically associated with the dependent variable; thus, plausible alternative causes can be eliminated (Colquitt, 2008). Researchers are then able to claim that the dependent variable is caused by the independent variable and that alternative causes can be ruled out. Consequently, experimental design provides a strong test of how robust a theory is (Zellmer-Bruhn, Caligiuri, & Thomas, 2016). If what is developed is an organizational theory, evidence of causality enables us to advise managers to implement practices grounded in theory with confidence, knowing that the intended outcomes can be expected to materialize (Tung & Stahl, 2018).

Over the last 10 years, the use of experimental methods has received an increasing level of attention in the field of Management and International Business. Editors of major journals have encouraged its use (see e.g. Colquitt, 2008; Reeb, Sakakibara, & Mahmood, 2012; van Witteloostuijn, 2015; Zellmer-Bruhn et al., 2016). However, experimental research is still under-represented in the field of International Management and is even rarer in language research. Having said this, the topic of language has been studied in Cross-cultural Psychology, Psycho-linguistics, International Marketing and Cognitive Science, where experiments are

used frequently. The current experimental language research in International Management has been influenced by these disciplines.

The objective of this chapter is not to explain how to design experimental research *per se*. Readers can identify many books and articles that offer detailed explanations on how to conduct experimental research. Instead, we first review the present use of experimental design in language research and then provide some explanations as to why experiments are rarely used. Using this information as the *baseline* for the present application of experimental methods in language research, we then illustrate how this method can be applied more broadly to advance research in the field.

Exploring the baseline: Experimental design in current language research

To establish the *baseline* for the present use of experiments in language research, we reviewed all papers included in two recent review articles (Karhunen et al., 2018; Tenzer et al., 2017). We found that although language researchers have used a large variety of research methods, including case studies, interviews, ethnographic studies, surveys, and conceptual research, there were only seven articles that used an experimental design. We subsequently identified a further article, published in 2017, that had used experimental methods. All eight articles are summarized in Table 1.

Using these eight articles as our baseline, we have identified a number of key characteristics of experimental research in the extant language in International Management literature.

- 1) *Outcome variables*. All eight articles focus on the same language effect, namely how the use of language (foreign vs. native language) affects people's attitudes (Harzing & Maznevski, 2002; Harzing et al. 2005; Harzing et al., 2009), decisions (Zander et al. 2011); individual behaviors, (Akkermans, Harzing & van Witteloostuijn, 2010; Urbig, Terjesen, Procher, Muehlfeld & van Witteloostuijn, 2016; Gargalianou, Urbig & van Witteloostuijn, 2017) and team behaviors (Comu, Unsal & Taylor, 2010). Researchers typically contextualized this language effect in the setting of corporate language use and language differences in culturally diverse teams and organizations. Furthermore, some researchers not only tested the language effect in isolation, but also investigated its interaction with culture (e.g. Zander et al, 2011).
- 2) *External validity*. A strong finding should have both internal validity (the effect can be attributed to the cause instead of other factors) and external validity (the effect can be generalized). Following the International Business research tradition, these articles demonstrate an international focus through their sample choices and the scale of their research. Four articles conducted large-scale cross-national studies in 7-24 countries (see Table 1). This multi-country sampling strategy enabled the researchers to test their focal language effects in different national and cultural contexts and using different languages. This enhanced the generalizability (i.e. external validity) of their findings.

Table 1: A list of articles using the experimental design in language research					
Authors, Year & Journals	Research Questions	Theories	Manipulations	Sample	
1. Harzing & Maznevski (2002), in <i>Language and Intercultural Communication</i>	Do respondents adjust their responses in a way that reflects the cultural values associated with the language of the questionnaire?	Cultural accommodation	Language of questionnaires (native vs. English)	Undergraduate students in 7 countries	
2. Harzing et al. (2005), in <i>International Journal of Crosscultural Management</i>	Does the use of English-language questionnaires lead to homogenization of responses?	Cultural accommodation	Language of questionnaires (native vs. English)	Undergraduate students in 24 countries [includes the 7 countries in #1]	
3. Harzing et al. (2009), in <i>International Business Review</i>	How does language and culture affect survey response styles (in both rating and ranking options)?	Cultural accommodation Linguistic confidence	Language of questionnaires and scenarios (native vs. English)	a) Undergraduate students in 24 countries [as in #2]; b) MBA students in 16 countries	
4. Akkermans, Harzing, & van Witteloostuijn (2010), in <i>Management International Review</i>	How does language (English vs. native) and cultural experience affect cooperative vs competitive behaviour?	Cultural accommodation	Language used in the games (native vs. English)	Undergraduate students in the Netherlands	
5. Comu, Unsal, & Taylor (2010), in <i>Journal of Management in Engineering</i>	How do cultural differences and language skills affect project network performance?	Cultural diversity and linguistic competence, transaction cost analysis	Monocultural teams (native English) vs. multicultural teams (English as 2 nd language)	University students [UG and PG] in USA	
6. Zander, Mockaitis, Harzing et al. (2011), in <i>Journal of World Business</i>	How do language (English vs. native) and culture affect managers' decisions (leadership related decisions)	Cultural accommodation	Language of scenarios and questionnaire (native vs. English)	MBA students in 17 countries [as in #3]	
7. Urbig, Terjesen, Procher, Muehlfeld & van Witteloostuijn (2016), in <i>Academy of Management Learning & Education</i>	Does the use of a native or a foreign language in the classroom impact students' propensity to free ride?	Dual process theory of higher cognition	Language used in the games (native vs. English)	Undergraduate students in the Netherlands	
8. Gargalianou, Urbig, & van Witteloostuijn (2017), in <i>Cross Cultural & Strategic Management</i>	How does using foreign languages affect cooperative behaviors in a prisoner's dilemma setting?	Cultural accommodation	Language used in the games (native vs. foreign)	Undergraduate students in Belgium	

- 3) *Experimental design and manipulation techniques.* Research design and manipulation techniques are essential elements of experimental research. Although the total number of articles using experimental designs is small, they have applied quite different types of designs with settings including lab environments (e.g. Urbig et al, 2016), the participants' natural environment (e.g. Harzing et al. 2009), and games (e.g. Gargalianou, Urbig & van Witteloostuijn, 2017). Additionally, these papers have explored a selection of manipulation techniques, including priming (e.g. Harzing & Maznevski, 2002) and vignettes (e.g. Zander et al. 2011).
- 4) *The use of student samples.* All eight studies used undergraduate students as their research sample, except for Zander et al. (2011), which used MBA students. Student research samples are very common in experimental design, especially in Economics and Psychology. It is a convenient and practical sample design choice. First, when the design includes computer simulation or other lab-based manipulations, students are much more accessible; it is difficult to recruit employees to participate in lab-based research. Further, in cross-country research, student samples are relatively homogenous compared with employee samples, so it is easier to ensure the comparability of results. Of course, student samples raise a potential challenge to internal validity. In the later part of the chapter, we discuss the pros and cons of using students as a sample and its implications for International Management and language research.

The low incidence of experimental research demonstrates that the baseline of experiments in language research is rather low. Moreover, in spite of the above-mentioned achievements, we have also identified considerable limitations in current experimental research in this field. First, only a very small number of researchers have engaged with this research method, with five out of the eight articles co-authored by Harzing and three co-authored by van Witteloostuijn, who has been advocating the use of experiments in International Business (van Witteloostuijn, 2015). Second, most of the articles were published between 2002 and 2011 and there has been a relative lack of activity in recent years. Finally, the majority of studies tested the cultural accommodation effect, which – building on the psycho-linguistics literature – tries to establish the extent to which using a foreign language leads participants to accommodate their attitudes and behaviors to the values of the culture associated with the foreign language in question. However, language research in IM has made considerable progress in recent years and many more language phenomena, effects, concepts, and mechanisms have been identified. Furthermore, psycho-linguistics is only one of the related research streams that we can draw upon; other research streams, such as cross-cultural psychology, general management and international marketing are also good sources of theories, mechanisms, and research design techniques. Therefore, there are still considerable unexplored opportunities for the use of experimental methods in the IM context. In the next sections, we discuss why we believe experimental research has not been widely used in the IB/IM field to date and illustrate how language researchers could explore this research method further.

Why are experimental designs rarely used?

The limited number of publications using experiments in language research does not necessarily imply that researchers are not interested in this method. However, they face a number of challenges that might make them reluctant to adopt experimental methods.

IB/IM research tradition favours survey research

Survey research is the dominant research method used in the IB/IM field at large. Survey research enables the researcher to test models with multiple antecedents, mediators, moderators and dependent variables in one study. This allows the researcher to capture the inherent complexity of IB/IM phenomena. In contrast, the majority of experimental research tests the effect of a small number of factors, normally one to three, in a smaller model. As a result, authors conducting experimental research might encounter reviewer concerns that other factors that might affect the dependent variable were not included in the model, making it potentially more difficult to publish experimental research.

We argue that this tradition might in fact be detrimental to the field of IB/IM. Although survey research has made significant contributions to our understanding of IB phenomena, like any research method it has its limitations. Survey research only generates evidence for correlational relationships between multiple antecedences and outcomes. It is also unable to rule out the possibility that the outcomes might be caused by unmeasured factors. As a result, although multiple factors can be included in a model, researchers cannot be entirely sure that the outcomes were indeed caused by these factors. In contrast, well-designed experiments display high levels of internal validity and can establish causality (Aguinis & Bradley, 2014). Although one experiment cannot identify all possible causes of a phenomenon, multiple studies can achieve this goal incrementally and, thus, ensure external validity. Both surveys and experiments have their strengths and weakness, but in order to understand IB/IM phenomena we need the contribution of both these research methods.

Lack of rigorous training and experience in experimental design

Many IB/IM researchers have considerable knowledge of and experience of conducting survey research, but are unfamiliar with experiments. They might be reluctant to change their research trajectory and use a research method with which they are not familiar. Good experiments require a rigorous research design. A flawed design can easily be picked up in the review process and these flaws are difficult to fix.

Moreover, experiments often test a single, clearly defined, relationship in one study. If an experiment is unsuccessful or produces null results, which may happen quite often, the data is unlikely to be publishable. This means that the time and resources that researchers have invested do not lead to a tangible outcome. In contrast, researchers have more flexibility when using survey methods, because they can test multiple hypotheses in one and the same study. Even if some of

the hypotheses are not supported, it is still possible to publish a paper. Facing institutional pressures for publication, IB/IM researchers may, thus, be inclined to stay with the safer option of survey research.

The developmental stage of language research

Language research in the IB/IM field is a relatively new area of research. Researchers have identified many theoretical mechanisms through interviews, case studies and ethnographic research, as well as conceptual work in the past two decades. To further theorize some situated, context-sensitive theories that have been developed through inductive research, the next step is to conduct deductive research to test specific hypotheses. Evidence of internal validity is needed so that the causal relationships within these theories can be conclusively demonstrated. Evidence of external validity is also needed so that the generalizability of these theories across contexts can be established. Quantitative research can thus enhance the current, predominantly qualitative, language research.

Experimental research in particular can make a valuable contribution in this respect. However, experiments are situated within a positivist philosophical research tradition, assuming the existence of an objective truth. They also involve a deductive approach. Its key research concerns are to test and refine existing theories through evidence of causality and generalizability (Corley & Gioia, 2011). To achieve these goals, the theories and mechanisms to be tested in experiments need to be conceptually well-established and empirically operationalized (or measured). However, many findings in language research have yet to be maturely operationalized; this raises challenges for experimental design. As a result, current experimental research has only used a narrow range of theories, in particular theories that have been well operationalized in other disciplines such as psycho-linguistics and cognitive science.

Moving beyond the baseline: Attitudes, design choices and illustrations for language research

Thinking experimentally

To facilitate the use of experimental designs, the first step is to develop a positive attitude toward experiments. Researchers need to be prepared to follow the call from the editors of *Journal of International Business Studies* to start “thinking experimentally” (Zellmer-Bruhn et al., 2016, p.404). Thinking experimentally means that we prepare ourselves to use this research method effectively. It involves accumulating knowledge about experimental methods, actively following the development of experimental research in IB and general management research, and searching for mechanisms, theories, and manipulations in related disciplines that can be introduced into our context. Furthermore, thinking experimentally means that when planning a new language research project, we give experiments the same consideration as other research methods. This requires researchers to develop the ability to identify phenomena suitable for experiments, to

formulate research questions that are appropriate to experiments, and to build a team that has the relevant expertise.

Experiments are in fact a very suitable research method for language research. A large amount of language research involves micro-level phenomena. Some key topics include: corporate language policies, individuals' use of a foreign language, the influence of one person's language ability on other employees or team members, and language-related discrimination or stereotypes. The dependent variables are usually individuals' cognition, perceptions, attitudes, behaviours, or decisions. Experimental design is appropriate to test the causal factors influencing these types of variables and might, thus, help us to advance current language research (van Witteloostuijn, 2015).

The remainder of this chapter outlines the key choices that need to be made in experimental design: student samples vs. mixed samples, between-subject vs. within-subject design, single vs. multiple studies, and manipulation techniques. Due to a lack of examples in existing language research, we have selected popular topics in language research and illustrate how experimental designs can be applied for each choice.

Design Choice 1: Student samples versus mixed sample designs

Experimental research is often criticized by Management scholars for its use of undergraduate student samples (Tung & Stahl, 2018), whereas in Psychology and Behavioural Economics student samples are perfectly acceptable. The key issue, however, is the extent to which students are representative of the target population. If the study in question investigates fundamental human processes, such as how native language vs. foreign language use affects people's cognition, using a student sample is legitimate, because students are a representative sample of the target population. Hence, the findings based on student data can be generalized across populations (Bello, Leung, Radebaugh, Tung, & Van Witteloostuijn, 2009).

In the IB field, however, the mechanisms of interest often reflect specific work contexts rather than basic human processes (Bello et al., 2009). As a result, students who have not worked in these specific work environments are unlikely to have the features of the target population that researchers want to investigate (Harrison & List, 2004). Some example of these specific contexts could be working overseas as expatriates or being subject to language policy decisions. If students do not have this type of experience, the manipulation might not trigger the hypothesized effect among student participants or trigger a response different from what was hypothesized. As a result, findings based on student samples might not be generalizable to the population of employees.

Illustration. In the context of language research, using a student sample could still have its merits. Although many students do not have work experience, they can have language experiences that resemble those of employees. For instance, in non-English speaking countries, an increasing number of universities have adopted an English-only policy. This policy might affect the universi-

ty's entry criteria, students' performance assessment and students' interactions with each other. This situation is similar to a corporate policy of using English as a corporate language and the consequences that may have within an organization. To some extent, students in these universities might thus have very similar experiences to employees working in organizations that use their non-native language as the corporate language, such as anxiety when using the language or being undervalued or marginalized due to low corporate language proficiency. Students could therefore be an appropriate example to study some of these basic psychological mechanisms in language research. However, students might not be a representative sample for certain interpersonal dynamics, because the interactions between students are different from those between employees who have defined roles and contractual relationships with the organization. In this situation, researchers can use a multi-sample design. For example, study 1 could use a student sample to test the hypothesis as a pilot study. If the hypothesis is supported, study 2 could replicate the result using real employees. Many web-based surveys (e.g. Qualtrics) and crowdsourcing tools (e.g. Amazon Mechanical Turk) can now help to reach target samples with specific characteristics (van Witteloostuijn, 2015). An ideal situation would be to then test the hypothesis in a real organizational setting in study 3 to eliminate possible self-selection bias in the web-based sample.

Design choice 2: Between-subject vs. within-subject design

The use of between-subject or within-subject design is an important choice when conducting experiments. Between-subject design requires different participants in the treatment and control groups. Each participant only receives *one* treatment. Normally, each group is advised to have a minimum of 20 cases (Simmons, Nelson, & Simonsohn, 2011). Thus, this type of design can be costly, as it requires different participants for treatment and control groups. Furthermore, in between-subject designs it is important to ensure that the experimental and control groups are comparable on any factors that might impact the dependent variable such as demographic characteristics. This can be achieved through randomization, i.e. randomly allocating participants to the treatment or control groups and can be checked using statistical procedures to compare groups. If these requirements are not met, then plausible alternatives cannot be ruled out. For example, if the randomization principle was not executed effectively, the mean age of the experimental group might be significantly different from that of the control group; thus, the observed effect could be caused by the age difference rather than the experimental manipulation.

A within-subject design asks participants to react to all treatments. Compared with between-subject design, this type of design requires a smaller number of participants. However, since every participant needs to react to multiple treatments, this design faces the risk of a carry over effect, meaning participants' reaction to a subsequent treatment might be affected by their prior treatments. This can be manifested through very different or very similar responses to conditions (Hsu, Simmons, & Wieland, 2017). If a strong carry over effect cannot be counter-balanced in the

design, then a within-subject design may not be a suitable choice. The carry over effect can be counter-balanced to some extent by randomizing the order of the treatments when presented to the participants. This is easily achieved through a computerized survey. In an article included in Table, 1, Zander et al. (2011) used a within-subject design. All participants were asked to respond to 6 different managerial decision-related scenarios (i.e. 1. rewarding individuals or teams, 2. CEO's decision-making, 3. manager's goal-setting; 4. face-saving, 5. conflict-resolution, 6. empathizing). The carry over effect was not an issue in this study because rather than focusing on the issues *per se*, the objective of the study was to identify whether the respondents' choices in these situations reflected patterns of the influence of their national cultural norms.

Illustration. The level of English language fluency and accents have been a popular topic in language research. One research angle could be accent-based discrimination. Researchers have found that a foreign language accent can affect the results of employment-related decisions (Hosoda & Stone-Romero, 2010) and evaluative judgments can be affected by people's foreign language accent (Hansen, Rakić & Steffens, 2014). This phenomenon can also be examined in the context of MNCs, such as how accent affects interactions between expatriates and host country employees. Using a between-subject design, researchers could prepare two audio clips, one with an expatriate speaking English without accent (the control condition) and the other with an expatriate speaking English with an accent, such as an Indian accent (the treatment condition). We could then randomly allocate employees (e.g. employees in the headquarters of an American MNC) to listen to one of the audio clips, and ask them to evaluate the ability of the speaker based purely on the audio information. This design could reveal how an English accent might bias people's evaluation of expatriates' ability. Of course, depending on the choice of accent and the nationality of expatriates and host country employees, the results may differ. Researchers can decide which choice to make based on their focal phenomenon and research question. A within-subject design is not suitable for this example because the two audio clips only vary in the accent and the content is kept constant. If participants are asked to listen to both clips, the carry over effect is likely to be strong. Additionally, after listening to both versions of the audio clip, participants are likely to guess the researcher's intention, and this understanding is likely to affect their subsequent evaluations.

Design choice 3: Single versus multi-study design

Although a single-study design can demonstrate the effect of the key mechanism, it cannot claim that the results are conclusive, whereas in a multi-study design, researchers can provide evidence with stronger internal and/or external validity. For example, researchers can test the main effect in one study, and then introduce a new variable (e.g. a moderator) in a follow-up study. In this way the design can test the main effect as well as its boundary conditions. Researchers could also test the effects of the focal variable on different dependent variables or its effect across different samples. Moreover, when testing the effect of a new variable, researchers can use different ways

to operationalize the same variable, such as by manipulating it in one study and measuring it using a scale in another study. In Fan and Harzing (2017), the researchers tested the effect of ethnic identity confirmation, a perceptual agreement between two people regarding the importance of their shared ethnicity. In one study, the focal concept was measured using a Likert scale; in another study it was manipulated using vignettes. The combination of the two approaches measuring the same construct – and providing very similar results – strengthened the internal validity of the research (Fan & Harzing, 2017). Finally, a multi-study design can also be a hybrid research design, such as mixing a lab experiment with field research (e.g. surveys). Lab experiments provide greater control and precision and provide evidence for causality. Survey research using real employees can contribute to generalizability (McGrath, 1981). Hence, a combination of both methods would create a much stronger study.

Illustration. Language researchers have revealed that using a corporate language policy affects the perceived status of employees in MNCs (Neeley & Dumas, 2016). To further test the relationship between language policy and perceived status, a multi-study experimental design can be deployed. The first study can be designed to establish the main effect of the introduction of a corporate language policy on employees' perceived status within the organization; the second study can replicate the effect using a different dependent variable such as perceived career advancement or intention to leave the organization; the third study could introduce another variable, such as employees' personal characteristics (e.g. language proficiency or nationality) and test how these variables affect the relationship between corporate language policy and perceived status. Furthermore, a survey study could be used to test whether the language policy influences actual or perceived status.

Design choice 4: Manipulation techniques

As illustrated in our discussion of current experimental research in the language in the IM field above, there are different manipulation techniques that can be used. Here we discuss three key options: language priming, vignettes and naturally occurring events.

Option 1: Language priming. According to social cognition research, our stored knowledge, such as stereotypes, goals or implicit theories, can be activated by contextual cues. Priming involves using specific information to activate this knowledge and examine how it influences subsequent social judgements and behaviors (Molden, 2014). For example, bicultural individuals are believed to have two sets of cultural knowledge that can be activated by contextual cues (Hong, Morris, Chiu, & Benet-Martínez, 2000). When studying American-Chinese biculturals, researchers used culturally prototypical images (e.g. Chinese flag and historical images vs. American flag and images) to activate the Chinese or American cultural knowledge set in participants' minds and tested whether participants' attitudes, judgments and behaviors were affected by the cultural knowledge thus activated (Cheng, Lee, & Benet-Martínez, 2006; Hong, Wan, No, & Chiu, 2007).

Illustration. Language can be used as a prime. It can activate cultural knowledge or country-related stereotypes, which can subsequently affect people's attitudes, judgements or behaviors. In some of the studies presented in Table 1, researchers used the survey language as a prime in traditional survey-based research, randomly distributing different language versions of the survey to participants. A comparison of the results under the two different conditions revealed the effect of language on participants' attitudes and behaviors. Similarly, researchers can also embed a language prime in well-established experimental games, such as the prisoner's dilemma game (Keysar, Hayakawa, & An, 2012), or simulated tasks (e.g. Comu et al, 2010).

Option 2: vignettes. Vignettes are "carefully constructed and realistic scenarios to assess dependent variables including intentions, attitudes, and behaviours" (Aguinis & Bradley, 2014, p. 352). A well-designed vignette can realistically present organizational reality and helps participants put themselves in this situation, even if they might not have personally experienced it. This type of manipulation enhances experimental realism and subsequently external validity (Aguinis & Bradley, 2014). By presenting vignettes to participants, researchers can solicit participants' responses to hypothetical situations. Vignettes can take different forms. In addition to a written format, researchers can also use pictures, audio or video vignettes. When audio or video is used, it is advisable to hire professionals with trained voices and trained performing skills in the production, so that the highest quality can be ensured. Vignettes are especially suitable for the study of sensitive topics (Aguinis & Bradley, 2014). Some topics might be unethical to study in a survey and/or could invite responses that reflect social desirability. In contrast, participants might feel more comfortable to provide honest responses when reacting to a hypothetical situation.

Illustration. A written vignette is relatively easy to develop and use. Among the eight papers we discussed above, the design by Zander and her colleagues (2011) used a combination of language manipulation (English vs. native) and scenarios of managerial decisions. They generated six managerial decision scenarios and presented them to participants (see Figure 1). They first gave participants a new identity, i.e. "a manager of a product division" and a brief description of the organizational context, i.e. "a product division that includes several workplace teams." This information set the parameters for the hypothetical situation. They then solicited participants' decisions on the "best way to reward higher performing employees in this division". Since the research instruments were distributed in multiple countries, they were able to compare if the responses were affected by national cultures or language.

Furthermore, with regard to language research, audio vignettes could be a very good choice. For example, researchers can manipulate accents or language proficiency through audio. If researchers plan to introduce the literature on leadership or emotions to language research, videos would be appropriate to manipulate leadership styles or emotions (e.g. anger). Finally, vignettes can be developed when researchers aim to investigate sensitive topics in language research. For example, asking employees to report their behaviors related to discrimination and stereotyping

against employees with lower corporate language ability in a survey could be too sensitive. However, asking them to report what they are likely to do based on a hypothetical situation might be more acceptable.

Figure 1: an example of a vignette in language research (vignettes were randomly distributed in the native language or English) (Zander, et al. 2011)

Scenario 1: Rewarding
You are a manager of a product division that includes several workplace teams. In your opinion what would be the best way to reward high performing employees in this division?
S1-1. Individual financial incentive based on each employee's individual performance (individual reward)
S1-2. A group-based financial incentive based on the results of the team (group reward)
S1-3. A profit-sharing scheme for all employees based on the performance of the entire company (profit sharing)
S1-4. Non-financial individual incentives (individual non-financial)
S1-5. Public recognition of the best performing employees (individual recognition)
S1-6. Public recognition of the best performing teams (team recognition)
S1-7. Faster promotion for high performing individuals (individual promotion)

Option 3: Naturally occurring events. Another type of manipulation (or treatment) is using natural events. Of course, we cannot actually manipulate a natural event. This type of experiment uses a naturally occurring event in the field as an experimental treatment and compares selected variables before and after the event (Harrison & List, 2004). We present two examples illustrating how these experiments can be designed. Example 1 used an unexpected event as a treatment, whereas Example 2 is a good example of how to turn a planned event into an experimental treatment.

Example 1: Olivas-Lujan, Harzing & McCoy (2004) tested whether the September 11 terrorist attacks, an unexpected event, changed the norms and values of American university students by combining two datasets originally collected for other projects. Data used in the study 1 of this article was collected from two American universities, namely the University of Virginia and the University of Pittsburgh. Although this was not part of the original design, the data in Virginia happened to be collected before September 11 (in April 2001), while data in Pittsburgh was collected after September 11 (in December 2001). Another wave of data from the same universities was collected late March/early April 2002. This unexpected event might have changed students' responses; thus, it created an opportunity for researchers to compare the data before and after the event. The authors found that the terrorist attacks on September 11 decreased the level of cosmopolitanism displayed by American students and increased their perceived importance of relationship hierarchy (Kluckhohn & Strodtbeck, 1961). In addition, they found that students'

ideal type of job had changed, with security of employment and having a job that allows one to serve one's country becoming significantly more important. These findings were supported for both the immediate effect and the delayed effect (Olivas-Luján, Harzing, & McCoy, 2011). This article used a second – completely unrelated – study to independently verify the results of the first study. Data for the pilot in study 2 were collected before September 11 (in March 2001), whereas data for the main study were collected after September 11 (in October 2001). A comparison of the data before and after the event supported the fact that September 11 increased the perceived importance of hierarchy, measured in this study by Hofstede's Power Distance dimension.

Example 2: Cheng and her colleagues (2011) conducted research to test whether the 2008 Beijing Olympics increased mainland Chinese' perceptions of Chinese and Western cultural differences. Thus, they investigated whether instead of promoting "One World one dream", holding the Olympics actually increased perceived inter-group differences. Study 1 of this research used a 2 (Time: before Olympics vs. after Olympics) X 2 (the Olympic salience: symbol exposure by adding a small Beijing Olympic symbol at the bottom right corner of each page of the questionnaire. vs. No symbol exposure) between subject design. The results revealed that Olympic salience had no effect on the perceived cultural differences in either type of set of values *before* the Olympics, but it increased the perceived cultural difference in values 3 months *after* the Games (Cheng et al., 2011). Study 2 of this research tested the effects of intercultural competitiveness priming through the Olympics experience in mainland China. This was a 3-factor between subject design with two manipulations: 2 (Time: before Olympics vs. after Olympics) X 2 (Brands: Chinese vs. American) and a measured factor (cultural identification). Different from study 1, this study included a control group. Student participants were asked to rate how strongly each Chinese (vs. American) brand evoked positive and negative intergroup emotions and how strongly they felt about each brand. Hong Kong students were used as a comparison group because although a part of China, Hong Kong enjoyed a high level of political autonomy and many Hong Kong Chinese do not perceive mainland Chinese as an ingroup. The study revealed that mainland participants expressed more positive emotions and had a more positive perception towards the mainland Chinese (vs. American) brand at the end of the Olympics, regardless of cultural identification. This result indicates that the Olympic experience strengthened ingroup favoring emotions and perceptions. In contrast, these patterns were not observed among participants from Hong Kong.

Since the experimental treatments used in these articles were real events (the September 11 attacks and the Beijing Olympics), they were able to offer greater realism than designs using artificial manipulations. Both examples described above used a pre-test and post-test design, so that the changes in the target variables could be observed. Example 1 used longitudinal data, making it possible to test if the effects of the event endured. If the natural event is a training program or a pilot launch of a new policy in an organizational context, evidence of a lasting effect can give organizations more confidence to implement the training program or new policy (Zellmer-Bruhn

et al., 2016). Example 2 included a control group that was not affected by the event using the same time points as the experimental group. By using a control group as a comparator, such as a “no-training” group or a unit of the organization not subjected to the new policy, an experimental design can eliminate plausible alternatives and increase robustness (Reeb et al., 2012).

Illustration: Many language-related events that occur in MNCs could be used as “natural events”, such as the implementation of a language-training program, the launch of a new language policy, or the arrival of a group of expatriates from the parent company of the MNC that changes the monolingual environment in the subsidiary. The key is to obtain information about planned events and make sure that key variables are measured before and after the event. It is also necessary to identify a control group that does not receive the treatment, and measure key variables in the control group roughly at the same time that they are measured in the treatment group. An example would be an organization launching a new language-training program. Researchers can measure target variables 3 months before and after the training program. They can also select employees who will not receive the training program as the control group. Understandably, it is not always possible for researchers to plan for naturally-occurring events. However, when an event happens, prepared researchers who “think experimentally” will be able to capture it and identify how an ad-hoc design can be implemented.

Conclusion

In this chapter, we have reviewed the current status of experimental research in language research in the international management literature. In its three decades of development, language research has mainly benefited from qualitative research and survey methods. To further advance our knowledge in this field, we need to test the causality of relationships and mechanisms that have been identified; this is where experimental designs can make a significant contribution. In this chapter, we have provided language researchers with many suggestions on designing effective experiments, such as using mixed samples (e.g. combining student and employee samples), mixed designs (e.g. combining experiments and surveys), vignettes and naturally occurring events. Given the current scarcity of experiments in the language research literature, there are ample opportunities for novel research with the use of experimental designs. We therefore issue a call to researchers to realize the tremendous potential of experiments in language research.

References

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational research methods, 17*(4), 351-371.

- Akkermans, D., Harzing, A.-W., & van Witteloostuijn, A. (2010). Cultural accommodation and language priming. *Management International Review*, 50(5), 559-583.
- Bello, D., Leung, K., Radebaugh, L., Tung, R. L., & van Witteloostuijn, A. (2009). From the editors: Student samples in international business research. *Journal of International Business studies*, 40(3), 361-364.
- Cheng, C.-Y., Lee, F., & Benet-Martínez, V. (2006). Assimilation and contrast effects in cultural frame switching: Bicultural Identity Integration and Valence of Cultural Cues. *Journal of Cross-Cultural Psychology*, 37(6), 742-760.
- Cheng, S. Y., Rosner, J. L., Chao, M. M., Peng, S., Chen, X., Li, Y., . . . Chiu, C. Y. (2011). One world, one dream? Intergroup consequences of the 2008 Beijing Olympics. *International Journal of Intercultural Relations*, 35(3), 296-306.
- Comu, S., Unsal, H. I., & Taylor, J. E. (2010). Dual impact of cultural and linguistic diversity on project network performance. *Journal of Management in Engineering*, 27(3), 179-187.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.
- Corley, K. G., & Gioia, D. A. (2011). Building theory about theory building: what constitutes a theoretical contribution? *Academy of Management Review*, 36(1), 12-32.
- Fan, S. X., & Harzing, A.-W. (2017). Host country employees' ethnic identity confirmation: Evidence from interactions with ethnically similar expatriates. *Journal of World Business*, 52(5), 640-652.
- Gargalianou, V., Urbig, D, & Witteloostuijn, A. v. (2017). Cooperating or competing in three languages: cultural accommodation or alienation?. *Cross Cultural & Strategic Management*, 24(1), 167-191.
- Hansen, K., Rakić, T., & Steffens, M. C. (2014). When actions speak louder than words: Preventing discrimination of nonstandard speakers. *Journal of Language and Social Psychology*, 33(1), 68-77.
- Harrison, G. W., & List, J. A. (2004). Field experiments *Journal of Economic Literature*, 42(4), 1009-1005.
- Harzing, A.-W. et al. (2005). Does the use of English-language questionnaires in cross-national research obscure national differences? *International Journal of Cross Cultural Management*, 5(2), 213-224.
- Harzing, A.-W. et al. (2009). Rating versus ranking: What is the best way to reduce response and language bias in cross-national research? *International Business Review*, 18(4), 417-432.

- Harzing, A.-W., & Maznevski, M. et al. (2002). The Interaction between language and culture: A test of the cultural accommodation hypothesis in seven countries. *Language and Intercultural Communication*, 2(2), 120-139.
- Hong, Y.-Y., Morris, M. W., Chiu, C.-Y., & Benet-Martínez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American Psychologist*, 55(7), 709-720.
- Hong, Y.-Y., Wan, C., No, S., & Chiu, C.-Y. (2007). Multicultural identities. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp. 323-346). New York Guilford Press.
- Hosoda, M., & Stone-Romero, E. (2010). The effects of foreign accents on employment-related decisions. *Journal of Managerial Psychology*, 25(2), 113-132.
- Hsu, D. K., Simmons, S. A., & Wieland, A. M. (2017). Designing entrepreneurship experiments. *Organizational research methods*, 20(3), 379-412.
- Karhunen, P., Kankaanranta, A., Louhiala-Salminen, L., & Piekkari, R. (2018). Let's Talk about Language: A Review of Language-Sensitive Research in International Management. *Journal of Management Studies*, 55(6), 980-1031.
- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The foreign-language effect: Thinking in a foreign tongue reduces decision biases. *Psychological science*, 23(6), 661-668.
- Kluckhohn, F. R., & Strodtbeck, F. L. (1961). *Variations in Value Orientations* Westport, Conn Greenwood Press.
- McGrath, J. E. (1981). Dilemmatics: The study of research choices and dilemmas. *American Behavioral Scientist*, 25(2), 179-210.
- Molden, D. C. (2014). Understanding priming effects in social psychology: What is "social priming" and how does it occur? *Social Cognition*, 32, (Supplement), 1-11.
- Neeley, T. B., & Dumas, T. L. (2016). Unearned status gain: Evidence from a global language mandate. *Academy of Management Journal*, 59(1), 14-43.
- Olivas-Luján, M. R., Harzing, A. W., & McCoy, S. (2011). September 11, 2001: Two quasi-experiments on the influence of threats on cultural values and cosmopolitanism. *International Journal of Cross Cultural Management*, 4(2), 211-228.
- Reeb, D., Sakakibara, M., & Mahmood, I. P. (2012). From the editors: Endogeneity in international business research. *Journal of International Business Studies*, 43, 211-218.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359-1366.

- Sutton, R., & Staw, B. (1995). What theory is not. *Administrative Science Quarterly*, 40(3), 211-218.
- Tenzer, H., Terjesen, S., & Harzing, A.-W. (2017). Language in International Business: A review and agenda for future research. *Management International Review*. 57(6), 815-854.
- Tung, R. L., & Stahl, G. K. (2018). The tortuous evolution of the role of culture in IB research: What we know, what we don't know, and where we are headed. *Journal of International Business Studies*, 49(9), 1167-1189.
- Urbig, D., Terjesen, S., Procher, V., Muehlfeld, K., & van Witteloostuijn, A. (2016). Come on and take a free ride: Contributing to public goods in native and foreign language settings. *Academy of Management Learning & Education*, 15(2), 268-286.
- Witteloostuijn, A. van (2015). Toward Experimental International Business. *Cross Cultural Management*, 22(4), 530-544.
- Zander, L. Mockatis, A. I.; Harzing, A.-W. et al. (2011). Standardization and contextualization: A study of language and leadership across 17 countries. *Journal of World Business*, 46(3), 296-304.
- Zellmer-Bruhn, M., Caligiuri, P. M., & Thomas, D. C. (2016). From the Editors: Experimental designs in international business research. *Journal of International Business Studies*, 47(4), 399-407.