An integrative review and theoretical framework of validity in qualitative research: Reflections on the Academy of Management Journal for 2000 to 2016

Won-Woo Park*
Hyemin Han**
Seon Hyeong Kim**
Soohyun Yoon***
Hyejin Yu****

〈Contents〉

I. Validity in Quantitative and Qualitative Research, and the Limitations
II. Application of a New Framework in Management
III. Discussion and Conclusion

Abstract

The prime objective of the current study is to establish a comprehensive framework for evaluating validity in qualitative research. Although utilizing qualitative research methodology warrants ample merits for generating new theories and understanding emerging phenomenon, a paucity of existing literature has

*Professor, College of Business Administration, Seoul National University.
**Doctoral student, College of Business Administration, Seoul National University.
***Doctoral student, Department of Management, Arizona State University, Tempe.
****Doctoral student, Department of Management, Oklahoma State University, Stillwater.

proposed an integrative framework to ensure the validity of qualitative findings. To fill this gap, we drew on the extant literature on qualitative research methodology and 195 Academy of Management Journal (AMJ) articles published during 2000-2016 and developed a new theoretical framework for evaluating validity. In particular, the current framework proposed six validity concepts in accordance to qualitative research stage—research design, data collection and data interpretation. Furthermore, to assess the progress made in qualitative research in management, we applied the proposed framework and analyzed 79 articles published in AMJ during 2012-2016. Our analyses showed that 3.54 validities were met on average, even though scholars' efforts to enhance validity have somewhat faded over the years. Consistent with the tradition of qualitative research emphasizing data richness, the category of operational measurement was fulfilled the most in general. Implications for theory and practice are discussed.

Keywords: Validity, Qualitative research, Research method

Qualitative research, which refers to the naturalistic approach that produces findings arrived from "real-world setting where the phenomenon of interest unfolds naturally" (Patton, 2002: 39), bears its greatest strength in its 'richness' (Weick, 2007). To illustrate, qualitative research allows an in-depth investigation of a contemporary phenomenon within its real-life contexts. Consequently, the credibility of the findings depends on the researcher, who becomes the instrument him/herself in the process (Patton, 2002). As the entire process of qualitative study is carried out through the subjective eye of researchers, one of the key weaknesses of qualitative research is validity (Gibbert & Ruigrok, 2010). Due to the lack of a standardized process for assessing validity, qualitative researchers often engage in 'persuasions' (Siggelkow, 2007) to ensure the validity trying to convince readers and reviewers about the credibility of methodological procedures and findings (e.g., Silverman, 2005). Therefore, qualitative research has been stigmatized by many to be a second-class research (Lee, Mitchell, & Sablynski, 1999). However, such stigmatization is highly unfortunate since qualitative research has been among the most insightful papers published in academia, generating novel theories (Eisenhardt, 1989).

With awareness of these misperceptions on qualitative research, many researchers have emphasized the establishment of validity in qualitative research, and thus, generated multiple

validity constructs from various perspectives (e.g., Cho & Trent, 2006; Golafshani, 2003; Maxwell, 1992; Winter, 2000). Accordingly, it intensified confusions due to the emergence of new terminologies and absence of consensus among scholars. To fill up the voids in previous discussion on validity, the purpose of this study is three-fold. First, we review previous methodological works on validity and present a theoretical framework that demonstrates how extant conceptualizations of validity were streamed. Second, based on the validity constructs in the past research and limitations inherited in it, we propose a well-organized framework that can thoroughly evaluate validity in qualitative research and related technics to enhance validity. Last, we review recent qualitative research in management and appraise their validity in light of our framework. With these purposes in mind, our study includes (1) the advancement of validity in quantitative and qualitative research, (2) three major streams of extant validity constructs in qualitative research, (3) limitations of previous frameworks, (4) an alternative guideline, and lastly, (5) reviews of validity in qualitative studies published in *Academy of Management Journal* based on an alternative framework. Discussions and conclusion will follow.

I. Validity in Quantitative and Qualitative Research, and the Limitations

1. Validity in Quantitative Research

Validity can be formally defined as "whether the means of measurement are accurate and whether they are actually measuring what they are intended to measure" (Winter, 2000: 3). Researchers have generally agreed with the notion that enhancing validity applies to all researchers regardless of which approach researchers may choose to employ either quantitative or qualitative (e.g., Golafshani, 2003; Bluhm, Harman, Lee, & Mitchell, 2011). Thus, researchers have long been concerned with the issue of validity to demonstrate the credibility of research findings. While the concept of validity is largely underdeveloped in qualitative research domains (Gibbert & Ruigrok, 2010), systematic procedures for assessing validity have been rigorously accumulated in quantitative research.

One important reason for the development of validity in quantitative research may arise from its distinct nature. Quantitative research is generally grounded on the positivist or scientific paradigm, with the aim of identifying the objective reality by measurements (Glesne & Peshkin, 1992). Thus, quantitative researchers emphasize accurate measurements and analyses of causal relationships between the studied variables (Denzin & Lincoln, 1998). Such emphasis on measurements and causal relationships led quantitative researchers to employ delicate experimental methods to test hypotheses (Hopefl, 1997). The importance of a rigorous research design in experimental settings and construct measurements took the notion of validity to the next level—more sophisticated and systematic agreements among scholars (Golafashni, 2003). Specifically, two separate flows in validity have been discussed in quantitative research. The first stream was the validity in research design and the other was in measurement (Kerlinger & Lee, 2000). Table 1 summarizes the validity constructs advocated by each stream.

Validity in research design proposed by Campbell and Stanley (1963), suggested two major components: internal and external validity. Internal validity, also known as interpretability,

〈Table 1〉 Validity Framework in Quantitative Research

<table>
<tr><td></td><td colspan="3">Validity in
Research Design</td><td colspan="3">Validity in
Measurement</td></tr>
<tr><td>Advocators</td><td>Campbell &
Stanley
(1963)</td><td></td><td>Cook & Campbell
(1979)</td><td>Joint committee
(1974)</td><td></td><td>Joint committee
(1984)</td></tr>
<tr><td rowspan="4">Validity
Framework</td><td rowspan="2">Internal validity</td><td rowspan="2">→</td><td>Internal validity</td><td>Content validity</td><td>→</td><td>Content−
related
validity</td></tr>
<tr><td>Statistical
Conclusion
validity</td><td>Concurrent validity</td><td rowspan="2">→</td><td rowspan="2">Criterion−
related
validity</td></tr>
<tr><td rowspan="2">External validity</td><td rowspan="2">→</td><td>External validity</td><td>Predictive validity</td></tr>
<tr><td>Construct validity</td><td>Construct validity</td><td>→</td><td>Construct−
related
validity</td></tr>
</table>

indicates appropriate observations on causal relationships between variables. On the other hand, external validity refers to generalizability, the extent to which findings are applicable to other contexts. Cook and Campbell (1979) later on refined this framework by adding statistical conclusion validity and construct validity. Another framework, validity in measurement, was devised by Joint Committee of three major associations—the American Psychological Association, the American Educational Research Association, and the National Council. Consisting of four-type validity: content, predictive, concurrent, and construct-validity, this framework mainly concerns the degree to which a study measures what it intends to measure (Kerlinger & Lee, 2000). By integrating the predictive validity and concurrent validity of the former framework into criterion-related validity, the joint committee later on established a three-type framework: content-related validity, criterion-related validity, and construct-related validity.

2. Validity in Qualitative Research

Compared to validity in quantitative research that reached a systematic framework, qualitative researchers described validity from a wide range of perspectives, thus generating multiple terms with similar meaning (Golafshani, 2003). Among numerous conceptualization and frameworks, extant literature on validity in qualitative research could be roughly classified into three streams: (1) identical terminologies with quantitative research (e.g., Campbell, 1975; De Massis & Kotlar, 2014; Eisenhardt, 1989; Lecompte & Goetz, 1982; Lincoln & Guba, 1985; Yin, 1984), (2) Unique frameworks reflecting the nature of qualitative research (e.g., Cho & Trent, 2006; Creswell, 1998; Eisner, 1991; Lather, 1993; Maxwell, 1992; Thomson, 2011), and (3) frameworks based on research stages (e.g., Maxwell, 1992; Yin, 1984). Table 2 shows the major streams of validity in qualitative research.

1) Identical terminologies with quantitative research

The first distinctive stream is to adopt identical validity terminologies with quantitative research (e.g., Campbell, 1975; De Massis & Kotlar, 2014; Eisenhardt, 1989; Lecompte & Goetz, 1982; Yin, 1984). For example, Lecompte and Goetz (1982) argued that the

〈Table 2〉 Review of Validity in Extant Qualitative Research

	Type	Examples	Terminologies and frameworks
1	Identical terminologies with quantitative research	Lecompte & Goetz (1982)	Internal validity, External validity, Reliability, and Objectivity
		Yin (1984)	Internal validity, External validity, Reliability, and Construct validity
		De Massis & Kotlar (2014)	Internal validity, External validity, and Construct validity
2	Unique terminologies reflecting the nature of qualitative research	Lincoln & Guba (1985)	Trustworthiness – Credibility, Transferability, Dependability, and Conformability
		Eisner (1991)	Structural corroboration, Consensual validation, and Referential adequacy
		Maxwell (1992)	Descriptive validity, Interpretive validity, Theoretical validity, Generalizability, and Evaluative validity
		Lather (1991, 1993)	Triangulation, Construct validity, Face validity, and Catalytic validity (1991) Ironic validity, Paralogic validity, Rhizomatic validity, and Voluptuous validity (1993)
		Creswell (1998)	Terminology of "validity" is not appropriated
		Cho & Trench (2006)	Transactional validity and Transformational validity
		Auerbach & Silverstein (2003), Thompson (2011)	Transparency
3	Frameworks Based on Research Stages	Yin (1984)	(1) External validity – stage of research design (2) Construct validity – stage of data collection and composition (3) Reliability – stage of data collection (4) Internal validity – stage of data analysis
		Maxwell (1992)	(1) Descriptive validity, Interpretive validity – early stages of data collection and interpretation (2) Theoretical validity, Generalizability – middle stages of data interpretation and abstraction process of the findings (3) Evaluative validity – last stage of the research

concepts of validity in quantitative research—internal validity, external validity, reliability, and objectivity—could be used to assess validity of qualitative research. Yin (1984) adopted the terminologies of validity from the quantitative research domain, such as construct validity, internal validity, external validity, and reliability. Although this stream was mainly prevalent at the early part of 1980's, it has subsided by the overflow of unique validity terminologies after the middle eighties. Recently, De Massis and Kotlar (2014) revived the assertion that qualitative research should be evaluated based on construct validity, internal validity, and external validity.

2) Unique terminologies reflecting the nature of qualitative research

Articulating the unique nature of qualitative research, some researchers have advocated using unique terminologies and meanings to conceptualize validity in qualitative research. Scholars in this stream argued that using the terminology of 'validity' is not appropriate to conceptualize credibility of qualitative research findings (e.g., Creswell, 1998), generating their own terms to illustrate validity in qualitative research. The initiation of unique terminologies was Lincoln and Guba's (1985) "trustworthiness"—credibility, transferability, dependability, and conformability; each of which is relevant to the concept internal validity, external validity, reliability, and objectivity. The formation of new terminologies continued throughout 1990's (e.g., Eisner, 1991; Lather, 1991; Maxwell, 1992, 1993). One notable and influential framework reflecting the trends in contemporary qualitative field is Cho and Trent's (2006). This framework suggested 'transactional validity'—the validity acquired through the active interactions between the researcher, the researched, and the collected data; and 'transformational validity'—the validity achieved by challenging and reconfirming the taken-for-granted meanings of prior qualitative research (Cho & Trent, 2006). This study put emphasis on unique aims of qualitative research—deeper understandings, broader visions, and novel insights or new implications (Lather, 1986, 1993; Richardson, 1997) and developed representative frameworks for qualitative researchers to consider essentially.

3) Frameworks based on research stages

Contrast to above two streams focusing on identical or different terminologies or meanings

with quantitative research, the last framework is driven by a more practical motivation. For example, Maxwell (1992) devised five types of validity—descriptive validity, interpretive validity, theoretical validity, generalizability, and evaluative validity; and allocated a series of validity to research stages. According to Maxwell's (1992) framework, descriptive validity and interpretive validity can be enhanced at the initial stage of research, theoretical validity and generalizability are related with the middle stages of research involving data collection and data interpretation, and lastly, evaluative validity is drawn at the last stage (Winter, 2000). Likewise, Yin (1984) suggested a validity framework based on research stages to make findings more validated. Although Yin's (1984) framework adopted the terms of validity in quantitative research—construct validity, internal validity, external validity, and reliability; it reflects the distinctive stages of case studies. For example, Yin (1984) suggested that external validity is relevant with the phase of research design, reliability with data collection, construct validity with data collection and composition, and internal validity with data analysis.

3. Limitations of Previous Studies on the Validity of Qualitative Research

Although scholars generally agree that validity in qualitative research cannot be a single and fixed concept, different frameworks and their inconsistent usages have been a big source of confusions (Atkinson et al., 2004; Winter, 2000). Based on literature review on qualitative research, this study points four main limitations inherited in previous discussions: (1) ambiguous terminologies and unorganized structures, (2) confusion of validity with reliability, (3) identical terminologies with quantitative research, and (4) incomplete coverage of the entire research stages. Detailed explanations about each point are as follows.

1) Ambiguous terminologies and unorganized structures

Unique validity terminologies of qualitative studies have been criticized because they are not clearly distinguished from each other. For example, Winter (2000) pointed out the ambiguity of Maxwell's (1992) five types of validity, mentioning that the meanings of five terminologies are intertwined and inseparable. This ambiguity issue is also embedded in many of other terminologies such as Eisner's (1991) structural corroboration, consensual validation;

Lather's (1991) triangulation, and Cho and Trent's (2006) transactional and transformational validity. In particular, the definitions of validity are more closely related to the techniques to enhance validity, rather than the concept itself. For example, Cho and Trent (2006) suggested that transactional validity can be acquired through the interactions between the research, the researched, and data, but did not elaborate which dimensions of validation could be increased by such interactions. Unorganized structures coming from unique terminologies are also problematic. Unlike quantitative research whose validity framework is well-established under the sub-classifications of research design and measurement (Cook & Campbell, 1979), extant validity frameworks in qualitative research are accumulated with non-systematic manners and most of them do not have any sub-classification scheme.

2) Confusion of validity with reliability

While validity and reliability are strictly differentiated in quantitative research, researchers in qualitative research domain argued that such distinction is not valid for qualitative approach. Rather, qualitative researchers stated that the demonstration of validity is sufficient to establish reliability (Lincoln & Guba, 1985; Patton, 2002). However, efforts for objectivity of research are required both in quantitative and qualitative research, and objectivity is clearly divided into validity and reliability (Kirk & Miller, 1986). While validity is related to the acceptability of relations between variables, reliability is related to the possibility of replication which is guaranteed by accurate data collections and measurements of the research (Kirk & Miller, 1986). Not only concepts, but also tactics to enhance validity and reliability can be distinguished. For example, discussions among researchers in data collection and coding process will enhance reliability since such activities are related to the accuracy of data. In contrast, making narratives by researchers in data analysis will improve validity because such activities are related to plausibility of findings. However, both concepts and tactics for validity have been mix-used with those for reliability in qualitative research. For example, descriptive validity proposed by Maxwell (1992) is closer to reliability because it refers to the extent to which data is accurate enough to reflect participants' responses.

3) Identical terminologies with quantitative research

Although criteria adopted from quantitative research domain have merit in evaluating the rigor of qualitative research, it bears limitations because these criteria do not fully consider the unique nature of qualitative research. Since the main goals of qualitative research are exploration of phenomena and theory building, rules of qualitative research are quite different from those of hypothetical deductive methods (Kirk & Miller, 1986). In addition, in quantitative research, controlling extraneous variables is very important as relationships among variables are hypothetically proposed (Kirk & Miller, 1986). In contrast, qualitative research does not propose hypotheses about the relationships among variables because its main purpose is to reveal such vague phenomena. Therefore, we believe that distinct validity concepts and meanings are needed in qualitative research.

4) Incomplete coverage of the entire research stages

According to Kirk and Miller (1986), there are four stages in qualitative research—invention, discovery, interpretation, and explanation; each of which is matched with research design, data collections and measurements, data analysis, and evaluation of findings. They argued that qualitative researchers should pay attention to validation required in each individual stage. In this vein, validity framework should cover the entire research stages with clearly distinguished terminologies in an exclusive and exhaustive manner. However, extant validity frameworks in qualitative research neither exhaustively nor exclusively cover all the research stages. For example, Maxwell's (1992) framework matched with the stages of qualitative research (Winter, 2000), but it only focuses on the stages of data collection and interpretation, missing research design stage. Although Yin (1984) covered all the research stages from research design to data analysis, it applies the same paradigm with quantitative research, and thus, does not fully elaborate exquisite aspects in deep-rooted in qualitative research.

4. A New Framework of Validity for Qualitative Research

To fill several limitations aforementioned, this study intends to establish a framework that

can ensure validity of qualitative research. To do so, we took several steps as follows. First, handbooks and articles dealing with validity concept in qualitative research were reviewed to obtain understanding of major terminologies, meanings and concepts used to explain validity (See Appendix A). Second, to investigate conditions under which researchers mainly enhanced validity, we reviewed previous qualitative studies, primarily focusing on methodology sections. Based on 195 qualitative research articles published in *Academy of Management Journal* from 2000 to 2017, we identified a full list of validity constructs that have been incorporated by researchers. Third, some validity concepts were excluded from our framework. In specific, ambiguous terminologies, notions primarily related to reliability, or concepts directly adopted from quantitative research domains were excluded (i.e. transactional validity, transformational validity, descriptive validity, dependability, internal validity, and external validity). Last, we extracted six validity concepts that were particularly highlighted in the prior *AMJ* publications.

Overall, the proposed framework in the current study distinguishes six validity concepts in accordance to qualitative research stage. According to Kirk and Miller (1986), qualitative research involves four stages that are mutually exclusive and exhaustive—invention (e.g., research design), discovery (e.g., data collection), interpretation (e.g., understanding), and explanation (e.g., communication)—all of which accompany distinctive research activities and outputs. Given that each research stage bears its unique purpose and activities respectively relevant to validity criteria (Maxell, 1992), organizing validity concept based on research stage warrants merit. For instance, Yin (1984) asserted that validity tests should be conducted throughout entire process of qualitative research process, rather than confining to research design. Morse, Barrett, Mayan, Olson, and Spiers (2002) also highlighted the need to increase the rigor of studies during the process of research, opposing the post-hoc evaluation in prior studies, especially conducting validity checking only at the phase of data analysis (Kirk & Miller, 1986; Lincoln & Guba, 1985). To note, Maxwell (1992) categorized five types of validity based on qualitative research stage. However, Maxwell's framework has often been a target of criticism for proposing redundant and overlapping validity concepts and thus being "paradoxical and unnecessary" (Winter, 2000). To complement for such limitation, this study proposes a conceptual framework that incorporates three key stages of qualitative research,

namely research design, data collection, and data analysis and interpretation. Furthermore, in line with prior emphasis, we posit that each research stage bears a unique criterion that needs to be evaluated and that researchers may strengthen the accuracy and validity of qualitative research by checking validities in each concrete step. The phase of explanation was not included in our framework since it is the stage of 'splitting-up with the scene' (Kirk & Miller, 1986), which is unlikely to be highly relevant with validity notion.

The benefit of tracing validity in the order of research process may be also beneficial to incorporate the diverse perspectives of people in research, such as participants, readers and reviewers. This is because strengthening validity during the entire process primarily requires more collaboration with participants and external parties, from research design (e.g., case selection) to data analysis (e.g., peer review). Furthermore, by providing detailed descriptions on the enhanced validity and deliberate tactics arranged by the order of research process, the entire process of qualitative research becomes more transparent. Transparency, which measures how well the researcher offers the detailed process of the study and thus enables the replication of research in similar environments (Auerbach & Silverstein, 2003), has recently been emphasized as to ensure whether the readers can clearly understand how the researchers could arrive at the final implication of the study (Bluhm et al., 2011). By providing explanation about which validity is enhanced in each stage and the specific technique used to do so, researchers can not only strengthen validity of the findings but also allow replication of the original study in the future research.

Drawing on the merits and demerits of validity based on research process, we refine previous validity categories and propose a validity framework according to three key phases of qualitative research: Research design, data collection, and interpretation. Each phase composes of two sub-categories by entailed research activities (See Table 3). Each research process is composed of two sub-categories, each of which proposes the validity that can be improved during the referred stage. Furthermore, we elaborate on the techniques that are relevant to each research process. The first stage of research is research design during which researchers can improve research question-case fit and generalizability; the second phase is data collection and researchers can enhance operational measurement and conceptual density; the last process is data interpretation during which researchers improve validity of concept and validity of

〈Table 3〉 Validity Framework Based on Research Process

	Research Process	Validity category	Definition	Supporting techniques
1	Research Design	Research question – case fit	Selecting research contexts to best serve research questions; transparently show the phenomenon in real world	Selecting extreme cases; Collaboration with participants
		Generalizability	Replication within the boundaries of research questions	Examining multiple cases; Mixed methods
2	Data Collection	Operational measurement	Acquiring ample amount of qualitative data to comprehensively cover the researched phenomena	Triangulation (source); Thick description; Purposive sampling
		Conceptual density	Refining emerging concepts and their relationships by exploring multiple categories of concepts	Disconfirming evidence; Snowball sampling; Collaboration; Iterative manner; Purposive sampling
3	Data Interpretation	Validity of concept	Enhancing the degree to which categories extracted from data represent the relevant narratives	Triangulation (investigator); Member check; Peer debriefing; Iterative manner
		Validity of pattern	Enhancing the degree to which the emergent theory can provide a cohesive story about the phenomenon in interest	Triangulation (investigator); Member check; Disconfirming evidence; Peer debriefing; Iterative manner

pattern.

5. Validity in Research Design

Building a research design involves the most initial stage of the overall research process. Making right decisions at this stage is pivotal because the validity of overall research cannot be secured if the chosen research design is not 'persuasive' enough. Hence, the main purpose

of building a research design is to avoid the situation in which the evidence does not address the proposed research questions (Yin, 1984). To do so, researchers need to carefully choose which research contexts or cases should be selected and thus investigated as to precisely address the proposed research questions. Given that research design is the 'the logical sequence' that connects the empirical data to not only a study's initial research questions but also to its ultimate conclusions (Yin, 1984), building a precise action plan during this stage bears a significant importance.

A review of prior qualitative research implies that securing validity at this research design stage primarily involves two phases: A proper fit between research question and the selected case, and generalizability of the potential qualitative findings. The former occurs at the most initial stage of building a research design since it is concerned with whether the chosen samples are appropriate or representative enough to yield an adequate answer to the research question. On the other hand, the latter can be contemplated at a comparatively later phase since it involves the extent to which the potential findings can be applied to similar groups or situations that share unique characteristics in common. Detailed explanations about validities that need to be considered at this stage are as follow.

1) Research question-case fit

Research question-case fit refers to selecting appropriate research contexts which can best serve research questions. Qualitative researchers need to choose their research context purposefully according to theoretically derived criteria. By providing a strong rationale for choosing a particular case, researchers can persuade why such research context can best answer and provide rich insights in accordance with research questions (Miles & Huberman, 1994).

Researchers may choose to *select extreme cases* to examine the phenomenon of theoretical interest. Extreme settings tend to make a phenomenon 'transparently observable' (Eisenhardt, 1989; Pettigrew, 1990) and thus issues of theoretical interest tend to be more visible and rich in its properties (Pratt, Rockmann, & Kaufmann, 2006; Weick, 2007). By choosing a context where phenomenon of interest unfolds obviously, researchers can secure research question–case fit to a more visible extent. Also, participants can be involved in selecting cases as co-researchers as well. For instance, Smith (2014) conducted a qualitative research to understand

how senior leaders embed paradoxes into their organization's strategy while also struggling to manage them effectively. To select cases for data collection, Smith (2014) collaborated with one of organization members, namely an executive in corporate strategy, to sample cases based on three criteria. Although not frequently employed, *collaborating with participants* at research design stage may be useful when researchers have difficulty getting fully informed of the research context at an early stage.

2) Generalizability

Generalizability signifies that the research outcomes derived from a specific context is applicable to different situations which share similar attributes with it. Generalizability of qualitative findings may be secured by employing two different techniques. The most commonly used procedure is to examine the research question across a multiple number of cases. This technique, namely *multiple-case approach*, adds robustness to the qualitative findings because multiple-case approach follows 'replication logics' (Eisenhardt, 1989; Yin, 1984). To illustrate, cases are chosen with an insight so that each can serve a specific purpose within the overall scope of inquiry. Each case either predicts similar results or produces contrary results but for predictions. Since such replication logic allows the researchers to gain confidence that findings are not attributed to sampling bias, multiple-case approach is the most commonly employed technique to enhance overall generalizability of qualitative research findings. Another way is to combine at least two methods, usually quantitative and qualitative, within a single research to address research questions (Morse, 1991). This approach, namely *mixed method* or *methodological triangulation*, provides the most comprehensive approach in solving the research problem by incorporating advantages of both methods (Morse, 1991). Thus, mixed method may be another way through which researchers can heighten generalizability of potential findings.

6. Validity in Data Collection

After confirming the overall research design, researchers embark on a full-scale qualitative data collection. Our review process revealed that data collection process comprises two

stages: Collecting multiple sources of data (e.g., data source triangulation) and taking iterative processes between data, theories, and the emerging constructs. In particular, researchers using the second method clarified that they adopted grounded theory analysis (Glaser & Strauss, 1967). Data collections using grounded theory analysis conduct data collections through several stages. This is because emerging concepts after initial data collection may make researchers find another evidence (i.e., data) to prove or disprove the looming construct (Strauss, 1987). In this section, we suggest two validity compositions that correspond with each sub-stages of data collection.

1) Operational measurement

Operational measurement refers to acquiring ample amount of qualitative data which are comprehensive enough to cover the researched phenomena. It should be considered at the first sub-stage of data collection process. The most representative way for researchers to enhance this validity is using *multiples sources of data*. For example, one of the common methods in qualitative research is to combine the case study utilizing one or more qualitative data sources with evidences from survey data. The depth and breadth of data that are acquired from various sources allow the researchers to have various perspectives on the research question. However, the richness of data per se is not the ultimate goal of acquiring data from multiple sources. The main goal of collecting data from multiple sources is the convergence of line of inquiries via triangulation (Denzin & Lincoln, 2011; Lincoln & Guba, 1985; Yin, 1979, 2013). Data collection from a single source bears a risk of the operational measurement being biased and may not fully convey the intended meanings of researchers. In this vein, Yin (1979) suggests an example of 'crime'. If researchers collect data only from the police reports to study crime, this study will rule out crimes that were not reported to police, and thus, end up partially measuring the concept of crime. Therefore, researchers obtain various data from multiple sources that captures the same phenomenon of interest to fully embody what researchers try to measure, and thus, enhance operational measurement. Thus, data collection stage is an essential groundwork for researchers to arrive at an objective conclusion by materializing the concept in an accurate and unbiased manner (Yin, 2013). In addition, researcher use *thick description* technique not only to amass detailed information

relevant to a particular episode but also to rightly interpret the intentions of participants in their words and actions (Holloway, 1997; Ponterotto, 2006; Schwandt, 2001). Furthermore, as phenomenon of qualitative interest cannot be readily captured by random sampling (Miles & Huberman, 1994), researchers also employ *purposive sampling* (Kuzel, 1992; Morse, 1989), which is one of the most representative tactics used in qualitative studies.

2) Conceptual density

In the above section, we examined that the breadth and depth of data lead to convergence of lines of inquiry, thereby alleviating potential biases in collected data and enhancing operational measurement. However, beyond the accurate and objective data collection, elaboration or refinement of theory generalization (Glaser & Strauss, 1967; Strauss,1987) has to be further achieved by taking note of conceptual density, which refers to refinement of emergent concepts and their relationships by testing whether the same findings are repeatedly observed with additional data from post-collection. The difference between the first sub-stage and this phase in data collection is that the former is the diversification of data sources aiming for accurate measurements of objects, whereas the latter is the diversification of samples intending the refinement or expansion of extant concepts or theories. In other words, the first enhances the 'accurateness' and 'objectivity' of the constructs through triangulation, while the second appreciate the 'density' of concept—the multiplicity of categories and properties and their relationships (Strauss, 1987), by collecting additional data or evidences.

During the theory-building process of inductive research, data collection is continued by moving iteratively, namely *iterative manner*, between the stages of coding (i.e., recording emerging constructs to conceptualize data) and previous arguments in literature to develop conceptual categories (Eisenhardt, 1989). The purpose of this technique is to make constant comparisons with the emerging concept. Collecting *disconfirming evidences* during these iterative processes is also a technic to validate emerging concepts and narratives of research. Also, researchers engage in *purposive sampling*, the intentional search for wider samples relevant to looming categories. Researchers may continue this purposive sampling until theoretical saturation is achieved—the additional data does not create any new aspects to the emerging categories (Glaser, 1978). Hence, the resultant categories after this

purposive sampling can construct denser conceptualization than initially emergent categories, contributing to sophistication of extant theories or concepts (Strauss, 1987). In addition, researchers can use *collaboration with participants* and *snowball sampling* technique identify and ensure appropriate data samples, and thus lead to denser conceptualization.

7. Validity in Data Analysis and Interpretation

The third stage of qualitative research is data analysis and interpretation. This stage aims to develop theoretical constructions that go beyond concrete descriptions about physical phenomena of interest (Maxwell, 1992). Generally, this stage consists of two steps grounded in constant comparative approaches (Glaser & Strauss, 1967). First, by engaging in extensive and iterative data analyses, researchers identify first order categories and convert them into higher-order themes. Second, researchers build theories that integrate categorical themes and offer coherent explanations about the researched phenomenon. The former step is often compared to brick-making operations, whereas the latter step is compared to the work of putting those bricks together (Maxwell, 1992). In line with prior arguments, we also propose that 'validity of concepts' and 'validity of patterns among concepts' should be correspondingly considered at the respective step to guarantee the credibility of qualitative research. The names of the validity are adopted from Maxwell (1992) who introduced these concepts as two distinct aspects of theoretical validity.

1) Validity of concept

Validity of concept refers to the degree to which the extracted categories represent the relevant narrative data (Maxwell, 1992). This validity is comparable to 'construct validity' from Kirk and Miller (1986) and is guaranteed by reaching a consensus on concepts within the research team. To enhance this validity, various tactics are used to resolve different views among the researchers and underpin the agreed-upon concepts. Among them, *triangulation by researcher* is the most representative technique. By making use of multiple views of investigators or theories, researchers can reduce the potential bias in data coding (Denzin, 1970) and corroborate evidences from different sources (Creswell, 1998). In practice, this

triangulation method is frequently aligned with *iterative manner* as investigators tend to independently code the concrete data, compare the results, and adjust final themes through discussions (e.g., Besharov, 2014; Powell & Backer, 2014; Lockett, Currie, Finn, Martin, & Waring, 2014). Furthermore, *peer debriefing* is another way to increase the validity by having external checks from other experts or colleagues in academia (e.g. Sonenshein, 2014; Smith, 2014). The role taken by external colleagues during peer-debriefing is often described as 'devil's advocates' as they ask hard questions about methods and interpretations used in the research, making research teams stay alert and honest. Last but not least, researchers can enhance the validity by getting feedbacks from participants of the study, namely *member check* procedure. Lincoln and Guba (1985: 314) highlighted this technique as "the most critical technique for establishing credibility" (e.g., Sonenshein, 2014; Besharov, 2014; Vaccaro & Palazzo, 2015).

2) Validity of patterns

Validity of patterns among concepts refers to the degree to which the emergent theory can provide a cohesive story about the phenomenon of interest (Maxwell, 1992). This validity becomes critical as researchers start to identify major story lines and write a story that can integrate categories extracted at the prior step. It is comparable to internal or causal validity proposed by Cook and Campbell (1979). Alike validity of concept, reaching an agreement within the researcher team is an essential factor for this validity of patterns. Therefore, techniques mentioned earlier, such as *triangulation by researchers*, *member check*, *iterative manner*, or *peer-debriefing*, are also commonly used at this second step. However, it should be noted that *searching for disconfirming evidences* is an additional technique that is applied in this latter step. It refers to a set of process where investigators search for evidence that either supports or disconfirms the preliminary themes or categories emerged during the initial data analysis (Creswell & Miller, 2000). As researchers are more inclined to find confirming evidence than disconfirming one, searching for disconfirming evidences is not an easy task. However, it can enhance the validity of the narratives by adding the sense of reality which is inherently complex and diverse (Creswell & Miller, 2000).

II. Application of a New Framework in Management

To confirm how well the above framework explains validity of qualitative research and identify which dimensions have been mostly addressed, we analyzed 79 qualitative studies published in *Academy of Management Journal* from 2012 to 2016 by applying the above six criteria based on research stages. Based on descriptions on techniques and intentions why such techniques were used in method sections, we determined whether each research meets our validity criteria.

1. Fulfillment of Validity Criteria in Qualitative Research

In 79 articles, about 3.54 validity criteria were met on average. Four of 79 articles fulfilled all six validity categories, while two articles fulfilled only one validity category. Operational

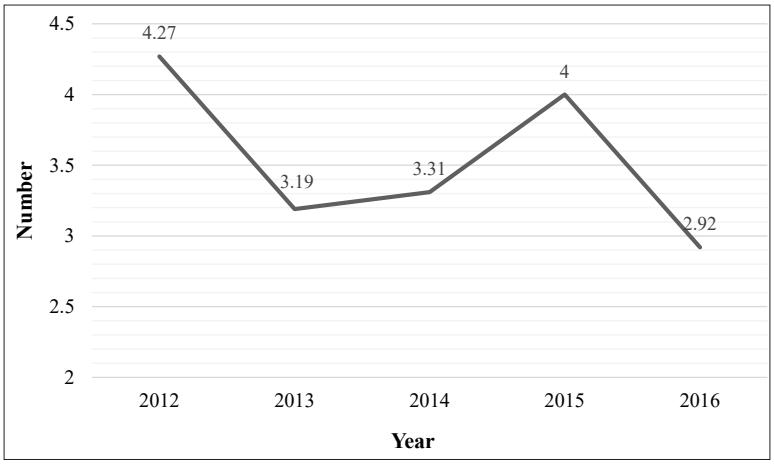
〈Table 4〉 Distribution in Number of Validity Criteria Fulfilled and Frequency of Each Validity Criteria

		Number of articles	Proportion (%)
Number of validity criteria	1	2	2.53
	2	19	24.05
	3	16	20.25
	4	22	27.85
	5	16	20.25
	6	4	5.06
	Total	79	100.00
Types of validity	Research question – case fit	55	69.62
	Generalizability	30	37.97
	Operational measurement	68	86.08
	Conceptual density	36	45.57
	Validity of concept	47	59.49
	Validity of pattern	44	55.70
	Total	79	100.00

measurement was the most frequently concerned validity in 79 studies. 68 of 79 (86.08%) were using at least one of tactics of triangulation (source), thick description, and purposive sampling in data collection. In contrast, generalizability was less emphasized relatively. Only 30 of 79 (37.97%) were using at least one of tactics of multiple cases and mixed methods in research design process. These results are consistent with the traditions of qualitative research domain which emphasized plentiful data collection and less concerned about applicability of research findings. Moreover, because some studies clearly mentioned that using multiple cases was not for enhancing applicability of findings of the study, generalizability was identified much less important type of validity for qualitative researchers. Other validity categories, such as research question–case fit (69.62%), conceptual density (45.57%), validity of concept (59.49%), and validity of pattern (55.70%), were fulfilled in greater or lesser degree.

2. Five–year Trend of Validity Fulfillment in Qualitative Research

As *AMJ* qualitative papers included in our reviews were published between 2012 and 2016, we tried to identify time-based trends in changes of the degree to which each validity type was considered. Although clear trends were not identifiable, we could take a glimpse that efforts to sustain validity in qualitative research has been weakening in recent years. Figure 1

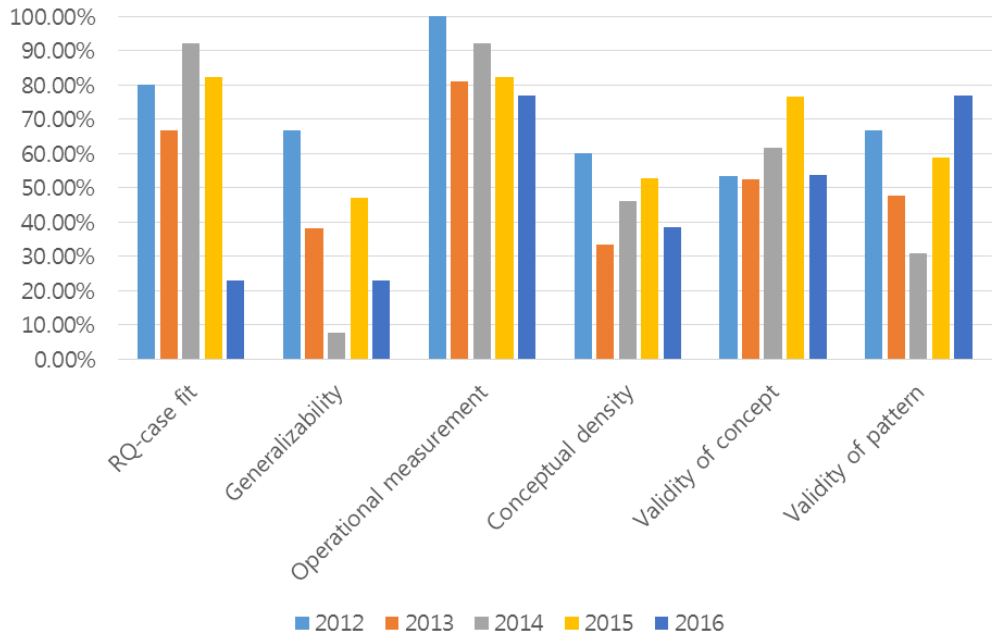


〈Figure 1〉 Five–year trend of validity fulfillment from 2012 to 2016.

summarize the changes of total number of validity categories fulfilled in the published articles for five years and Figure 2 present trends of changes in the proportion of articles which fulfilled each validity during the same period.

Figure 1 shows an up-and-down pattern in the average number of total validity categories which are fulfilled in the examined articles between 2012 and 2016. In 2012, the score took the highest point of 4.27 but the number declined to 3.19 and 3.31 in subsequent years of 2013 and 2014. In 2015, the number rebounded to the point of 4.00, but in 2016, it dropped again to the lowest point of 2.92, which is even smaller than half of the number of total validity categories. To sum up, it can be concluded that the qualitative researchers' efforts to enhance research validity has weakened over the years.

Likewise, Figure 2 shows similar fluctuations in the percentage of articles that fulfilled each validity categories between 2012 and 2016. For reference, the number of articles published each year is 15 for 2012, 21 for 2013, 13 for 2014, 17 for 2015, and 13 for 2016. Detailed explanations for each graph are as follows: First, generalizability, which is the least concerned validity, was emphasized the most in 2012 and nearly 70 percent of articles satisfied conditions for it. However, it has become of little accounts—with the exception of slight rebound in 2015—and only 23 percent of articles utilized strategies for generalizability. Second, the percentage of articles fulfilling research question–case fit remained relatively consistent—around 70 or 80 percent—during the first four years. However, a sharp decline emerged in 2016 and only 23 percent of articles adopted strategies to guarantee research question–case fit. Third, a majority of the examined articles took heed of operational measurement during the entire periods. In particular, more than 80 percent of articles fulfilled this validity during the first four years. This reflects the tradition of qualitative research that strongly emphasizes the data richness. However, the graph shows gradual declinations since 2013, and in 2016, approximately 77 percent of articles considered operational measurement. Fourth, the percentage of articles satisfying conceptual density remained relatively low during the entire period. This may reflect the fact that acquiring additional data is not an easy task. However, we could observe some variance among different years: The proportion of articles satisfying this validity recorded the highest point—60 percent—in 2012 and took the lowest point—33 percent—in 2013. In subsequent years of 2014 and 2015, the proportion increased



〈Figure 2〉 Distribution of each validity criteria per year.

gradually and in 2016, the graph showed slight decline which ends up at 38 percent. Fifth, validity of concept was considered the most in 2015 since 78 percent of articles utilized strategies to enhance this validity. However, it held in slight regard in the following year as nearly half of articles fulfilled the conditions for this validity. Last, validity of pattern showed V-curve during the analyzed period, taking the lowest point in 2013—30 percent—and recovering to the highest point in 2016—77 percent. By comparing the percentage of articles fulfilling respective validity each year, it becomes evident that a majority of validity categories—generalizability, research question–case fit, operational measurement, conceptual density, and validity of concept—have been regarded less in recent years. This calls for urgent needs to invest additional efforts to guarantee validity. As an exception, validity of pattern showed continuous increase in the percentage of articles fulfilling the validity condition between 2013 and 2016.

3. Reviews of Techniques in Qualitative Research

Subsequent to the reviews of validity fulfillment in published papers, we analyzed the

〈Table 5〉 Distribution of Each Validity Technique

Types of techniques	Number of articles	Proportion (%)
Multiple cases	20	25.32
Mixed method	13	16.46
Selecting extreme cases	55	69.62
Collaboration	2	2.53
Triangulation (source)	65	82.28
Thick description	10	12.66
Purposive sampling	6	7.59
Purposive sampling	26	32.91
Iterative manner	13	16.46
Snowballing sampling	5	6.33
Disconfirming evidence	3	3.80
Collaboration	2	2.53
Iterative manner	29	36.71
Triangulation (investigator)	19	24.05
Peer debriefing	8	10.13
Member check	5	6.33
Iterative manner	23	29.11
Member check	18	22.78
Triangulation (investigator)	13	16.46
Peer debriefing	9	11.39
Disconfirming evidence	3	3.80

frequency of strategies used to enhance individual validity. The results are summarized in Table 5. First, among ways to enhance generalizability, utilizing multiple cases was preferred

to mixed method. Second, as for research question–case fit, selecting extreme cases took an overwhelming proportion, whereas collaboration technique was hardly used, despite its effectiveness in selecting appropriate research contexts where researchers are less familiar with. Third, to guarantee operational measurement, triangulation by multiple sources was the most popular technique, followed by thick description and purposive sampling. Fourth, among conceptual density, purposive sampling exceeded the majority and was followed by iterative manner. Fifth, for validity of concept, iterative manner was the most popular strategy and triangulation by investigator was the second most popular approach. Confirmation by third parties was used less frequently for this validity. Lastly, among ways to enhance validity of patterns, iterative manner, member check, and triangulation by investigator were the three most frequently used techniques.

III. Discussion and Conclusion

1. Findings and Implications

Overall, this study offers three major contributions to the ongoing debate about criteria for differentiating a 'good' qualitative research from 'bad' qualitative research. First, this study sheds light on the unique characteristics of validity notion within qualitative research paradigm. Although several scholars have suggested that the notion of validity and techniques used to ensure it may differ depending on the research design employed (i.e., quantitative or qualitative), to our knowledge, this is the first study to highlight the unique quality of validity in qualitative research via providing systematic review of validity frameworks established in quantitative research method. As such, our comparative discussion warrants benefit to enrich our understanding of distinct characteristics of validity in qualitative study.

Second, the present study provides a critical review of prior discussion of validity in qualitative research and points out key limitations in the extant literature. Scholars have generally supported the notion that qualitative research method needs to ensure validity and suggested a number of validity constructs from various perspectives. Although qualitative

research has witnessed a notable progress from emerging validity concepts and corresponding techniques that may enhance specific validity, confusion has intensified at the same time due to absence of consensus among scholars. By critically reviewing multiple constructs suggested in the literature, the present study pointed out key limitations and suggested constructive directions as to define validity in qualitative research.

Third, this study offers an alternative validity framework as a practical checklist for scholars who desire to ensure validity when conducting qualitative research. Although several researchers have suggested frameworks for enhancing validity of qualitative research, these frameworks bear four significant limitations that we have addressed. To fill this gap, the current study suggested a framework that not only incorporates unique nature of qualitative study but also reflects the trend in reporting qualitative findings. Furthermore, we reviewed 79 qualitative studies published in *Academy of Management Journal* and proved merits of employing this framework to evaluate validity of empirical qualitative studies.

2. Limitations and Recommendations for Future Research

Several limitations within this study should be mentioned. First, we made extensive inquiries into qualitative research but did not specify each type of qualitative studies. Representative approach of qualitative research include ethnography, phenomenology, grounded theory, participatory action research, and case study and each type has distinctive study aims and data collection methods (Law, Stewart, Letts, Pollock, Bosch, & Westmorland, 1998). Thus, each approach may have different standards on validity. In connection with this limitation, this study derives a validity guideline around qualitative research in management literature, whose approach is mainly focused on case study. Therefore, our criticism and suggestions about qualitative research may not be applied in other sub-categories in qualitative research. Future research need to contextualize validity in terms of each method in qualitative study and refine our framework.

An additional limitation of this study is the lack of explanations on specific techniques to enhance validity. Although our framework suggests several strategies to improve validity-related traits, we neither covered detailed procedures nor strengths and weaknesses for each

technique. This is because our main concern of this study is to develop a framework based on the concept of validity itself. Future studies should fill this gap and suggest more practical guides that may lay out specific action plans around detailed techniques.

The issue of reliability should also be noted in future research. From quantitative perspective, reliability refers to the extent to which findings are consistent over a time within the same research setting. Although some prior studies have utilized identical reliability construct to evaluate qualitative research (e.g., Lecompte & Goetz, 1982; Yin, 1984), the matter of 'repeatability' in the context of qualitative approach has rarely been explored in depth. Can qualitative researchers guarantee the consistent findings in the same context if the study is conducted by other researchers? If not, can we say that the study is not rigorous? Or, should we say that it still holds its own significance? With regard to this matter, future studies need to devise a more refined lens to capture reliability in qualitative research.

3. Conclusion

The current study represents a meaningful contribution to our understanding of validity notion in qualitative research paradigm. By reviewing prior discussion of validity in qualitative research and proposing a new framework that can productively guide researchers to ensure validity, this study attempted to expand our methodological toolset for enhancing validity in qualitative research, particularly in management and organizational studies. Further theoretical progress and research in this area is critical not only to enrich our understanding of validity in qualitative research but also to generate qualitative studies of exceptional quality that can add more insightful theories to the field.

Appendix A

Authors	Title of handbook and article
Bluhm et al. (2011)	Qualitative research in management: A decade of progress
Brink (1993)	Validity and reliability in qualitative research
Cho & Trent (2006)	Validity in qualitative research revisited
Creswell (1998)	Qualitative inquiry and research design: Choosing among five traditions
Creswell & Miller (2000)	Determining validity in qualitative inquiry
De Massis & Kotlar (2014)	The case study method in family business research: Guidelines for qualitative scholarship
Eisenhardt (1989)	Building theories from case study research
Eisner (1991)	The enlightened eye: Qualitative inquiry and the enhancement of educational practice
Glasser & Strauss (1967)	The discovery of grounded theory
Healy & Perry (2000)	Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm
Kirk & Miller (1986)	Reliability and validity in qualitative research
Lather (1991)	Getting smart: Feminist research and pedagogy with/in the postmodern
Lather (1993)	Fertile obsession: Validity after poststructuralism
Lecompte & Goetz (1982)	Problems of reliability and validity in ethnographic research
Lincoln & Guba (1985)	Naturalistic inquiry
Maxwell (1992)	Understanding and validity in qualitative research
Savall et al. (2008)	The emergence of implicit criteria actually used by reviewers of qualitative research articles: Case of a European journal
Strauss (1987)	Qualitative analysis for social scientists
Thomson (2011)	Qualitative research: Validity
Yin (1981)	The case study crisis: Some answers
Yin (1984)	Case study research: Design and methods

References

American Association of College for Teacher Education. (2002). ***Governmental relations update***. Washington, DC: AACTE.

Amis, J. M., & Silk, M. L. (2008). The philosophy and politics in qualitative organizational research. ***Organizational Research Methods, 11***(3), 456-480.

Atkinson, P., Delamont, S., & Coffey, A. (2004). ***Key themes in qualitative research: Continuities and changes***. Walnut Creek, CA: Rowman AltaMira Press.

Auerbach, C., & Silverstein, L. B. (2003). ***Qualitative data: An introduction to coding and analysis***. New York, NY: New York University Press.

Besharov, M. L. (2014). The relational ecology of identification: How organizational identification emerges when individuals hold divergent values. ***Academy of Management Journal, 57***(5), 1485-1512.

Bingham, C. B., & Davis, J. P. (2012). Learning sequences: Their existence, effect, and evolution. ***Academy of Management Journal***, ***55***(3), 611-641.

Bluhm, D. J., Harman, W., Lee, T. W., & Mitchell, T. R. (2011). Qualitative research in management: A decade of progress. ***Journal of Management Studies, 48***(8), 1866-1891.

Brink, H. (1993). Validity and reliability in qualitative research. ***Curationis, 16***(2), 35-38.

Campbell, D. T. (1975). III. "Degrees of freedom" and the case study. ***Comparative Political Studies, 8***(2), 178-193.

Campbell, D. T., & Stanley, J. C. (1963). ***Experimental and quasi-experimental designs for research***. Boston, MA: Houghton Mifflin.

Cho, J., & Trent, A. (2006). Validity in qualitative research revisited. ***Qualitative Research, 6***(3), 319-340.

Cook, T. D., & Campbell, D. T. (1979). ***Quasi-experimentation: Design and analysis issues for field settings***. Boston, MA: Houghton Mifflin.

Creswell, J. W. (1998). ***Qualitative inquiry and research design: Choosing among five traditions***. Thousand Oaks, CA: Sage Publications.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. ***Theory into Practice, 39***(3), 124-130.

Davies, D., & Dodd, J. (2002). Qualitative research and the question of rigor. ***Qualitative Health Research, 12***(2), 279-289.

De Massis, A., & Kotlar, J. (2014). The case study method in family business research: Guidelines for qualitative scholarship. ***Journal of Family Business Strategy, 5***(1), 15-29.

Denzin, N. K. (1970). Strategies of multiple triangulation. In N. K. Denzin (Ed.), ***The research act in sociology: A theoretical introduction to sociological method (pp. 297***-313). New York: McGraw-Hill.

Denzin, N. K., & Lincoln, Y. S. (1998). ***Collecting and interpreting qualitative materials***. Thousand Oaks, CA: Sage Publications.

Denzin, N. K., & Lincoln, Y. S. (2011). ***The Sage handbook of qualitative research*** (4th ed.). Thousand Oaks, CA: Sage Publications.

Eisner, E. W. (1991). ***The enlightened eye: Qualitative inquiry and the enhancement of educational practice***. New York, NY: Macmillan.

Eisenhardt, K. M. (1989). Building theories from case study research. ***Academy of Management Review, 14***(4), 532-550.

Gibbert, M., & Ruigrok, W. (2010). The "what" and "how" of case study rigor: Three strategies based on published work. ***Organizational Research Methods, 13***(4), 710-737.

Glaser, B., & Strauss, A. (1967). ***The discovery of grounded theory***. Chicago, IL: Aldine Publishing Company.

Glesne, C., & Peshkin, P. (1992). ***Becoming qualitative researchers: An introduction***. New York, NY: Longman.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. ***The Qualitative Report, 8***(4), 597-606.

Healy, M., & Perry, C. (2000). Comprehensive criteria to judge validity and reliability of qualitative research within the realism paradigm. ***Qualitative Market Research: An International Journal, 3***(3), 118-126.

Hoepfl, M. C. (1997). Choosing qualitative research: A primer for technology education researchers. ***Journal of Technology Education, 9***(1), 47-63.

Holloway, I. (1997). ***Basic concepts for qualitative research***. Wiley-Blackwell.

Howard,-Grenville, J., Metzger, M. L., & Meyer, A. D. (2013). Rekindling the flame: Processes of identity resurrection. ***Academy of Management Journal***, ***56***(1), 113-136.

Johnson, B. R. (1997). Examining the validity structure of qualitative research. ***Education, 118***(3), 282-292.

Kerlinger, F. N., & Lee, H. B. (2000). ***Foundations of behavioral research*** (4th ed.). Belmont, CA: Cengage Learning.

Kirk, J., & Miller, M. L. (1986). ***Reliability and validity in qualitative research***. Beverly Hills, CA: Sage Publications.

Kuzel, A. J. (1992). Sampling in qualitative inquiry. In B. Crabtree & W. Miller (Eds.), ***Doing qualitative research: Research methods for primary care*** (pp. 31-44). London, England: Sage Publications.

Lather, P. (1986). Issues of validity in openly ideological research: Between a rock and a soft place. ***Interchange, 17***(4), 63-84.

Lather, P. (1991). ***Getting smart: Feminist research and pedagogy with/in the postmodern***. New York, NY: Routledge.

Lather, P. (1993). Fertile obsession: Validity after poststructuralism. ***The sociological quarterly, 34***(4), 673-693.

Law, M., Stewart, D., Letts, L., Pollock, N., Bosch, J., & Westmorland, M. (1998). Guidelines for critical review of qualitative studies. ***McMaster University Occupational Therapy Evidence-Based Practice Research Group***.

LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. ***Review of Educational Research, 52***(1), 31-60.

Lee, T. W., Mitchell, T. R., & Sablynski, C. J. (1999). Qualitative research in organizational and vocational psychology, 1979-1999. ***Journal of Vocational Behavior, 55***(2), 161-187.

Lincoln, Y. S., & Guba, E. G. (1985). ***Naturalistic inquiry***. Beverly Hills, CA: Sage Publications.

Lockett, A., Currie, G., Finn, R., Martin, G., & Waring, J. (2014). The influence of social position on sensemaking about organizational change. ***Academy of Management Journal, 57***(4), 1102-1129.

Mair, J., Marti, I., & Ventresca, M. J. (2012). Building inclusive markets in rural Bangladesh: How intermediaries work institutional voids. ***Academy of Management Journal, 55***(4), 819-850.

Maxwell, J. (1992). Understanding and validity in qualitative research. ***Harvard Educational Review, 62***(3), 279-301.

Miles, M. B., & Huberman, A. M. (1994). ***Qualitative data analysis: An expanded sourcebook*** (2nd ed.). Thousand Oaks, CA: Sage Publications.

Morse, J. M. (1989). Strategies for sampling. In J. M. Morse (Ed.), ***Qualitative nursing research: A contemporary dialogue*** (pp. 117-131). Rockville, MD: Aspen.

Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation.

Nursing Research, 40(2), 120-123.

Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. ***International Journal of Qualitative Methods, 1***(2), 13-22.

National Research Council. (2002). ***Scientific research in education***. Washington, DC: National Academy Press.

Patton, M. Q. (2002). ***Qualitative evaluation and research methods*** (3rd ed.). Thousand Oaks, CA: Sage Publications.

Pettigrew, A. (1990). Longitudinal field research on change: Theory and practice. ***Organization Science, 1***(3), 267-292.

Ponterotto, J. G. (2006). Brief note on the origins, evolution, and meaning of the qualitative research concept thick description. ***The Qualitative Report***, ***11***(3), 538-549.

Powell, E. E., & Baker, T. (2014). It's what you make of it: founder identity and enacting strategic responses to adversity. ***Academy of Management Journal, 57***(5), 1406-1433.

Pratt, M. G. (2008). Fitting oval pegs into round holes: Tensions in evaluating and publishing qualitative research in top-tier North American journals. ***Organizational Research Methods, 11***(3), 481-509.

Pratt, M. G., Rockmann, K. W., & Kaufmann, J. B. (2006). Constructing professional identity: The role of work and identity learning cycles in the customization of identity among medical residents. ***Academy of Management Journal, 49***(2), 235-262.

Richardson, L. (1997). ***Fields of play: Constructing an academic life***. New Brunswick, NJ: Rutgers University Press.

Savall, H., Bonnet, M., & Peron, M. (2008). The emergence of implicit criteria actually used by reviewers of qualitative research articles: Case of a European journal. ***Organizational Research Methods, 11***(3), 510-540.

Schwandt, T. A. (2001). ***Dictionary of qualitative inquiry*** (Vol. 31, pp. 439-448). Thousand Oaks, CA: Sage Publications.

Siggelkow, N. (2007). Persuasion with case studies. ***Academy of Management Journal, 50***(1), 20-24.

Silverman, D. (2005). ***Doing qualitative research: A practical handbook*** (2nd ed.). Thousand Oaks, CA: Sage Publications.

Smith, W. K. (2014). Dynamic decision making: A model of senior leaders managing strategic paradoxes. ***Academy of Management Journal, 57***(6), 1592-1623.

Sonenshein, S. (2014). How organizations foster the creative use of resources. ***Academy of Management Journal, 57***(3), 814-848.

Stenbacka, C. (2001). Qualitative research requires quality concepts of its own. ***Management Decision, 39***(7), 551-555.

Strauss, A. (1987). ***Qualitative analysis for social scientists***. Cambridge, England: Cambridge University Press.

Thomson, S. (2011). Qualitative research: Validity. ***Journal of Administration & Governance, 6***(1), 77-82.

Vaccaro, A., & Palazzo, G. (2015). Values against violence: Institutional change in societies dominated by organized crime. ***Academy of Management Journal, 58***(4), 1075-1101.

Walsh, K. (2003). Qualitative research: Advancing the science and practice of hospitality. ***The Cornell Hotel and Restaurant Administration Quarterly, 44***(2), 66-74.

Weick, K. (2007). The generative properties of richness. ***Academy of Management Journal, 50***(1), 14-19.

Winter, G. (2000). A comparative discussion of the notion of validity in qualitative and quantitative research. ***The Qualitative Report, 4***(3), 1-14.

Yin, R. K. (1979). ***Changing urban bureaucracies: How new practices become routinized***. Lexington, Mass: Lexington Books.

Yin, R. K. (1981). The case study crisis: Some answers. ***Administrative Science Quarterly, 26***(1), 58-65.

Yin, R. K. (1984). ***Case study research: Design and methods***. Beverly Hills, CA: Sage Publications.

Yin, R. K. (2013). ***Case study research: Design and methods*** (5th ed.). Thousand Oaks, CA: Sage Publications.

Zenbinati, V. S. S., & Liu, G. (2012). Which iron cage? Endo- and exoisomorphsim in corporate venture capital programs. ***Academy of Management Journal, 55***(2), 477-505.

질적 연구 타당성에 관한 비판적 고찰과 통합적 프레임워크 제안: 2000~2016년 *Academy of Management Journal*에 실린 질적 연구 논문들을 바탕으로

요 약

질적 연구는 기존에 연구가 덜 이루어졌던 현상을 탐색함으로써 새로운 이론을 제안하는 데 유용한 연구방법이다. 하지만 질적 연구는 학자마다 연구의 타당성 충족 요건을 다르게 제시하고 있어, 타당성 평가의 객관적인 기준이 될 통합적 프레임워크가 부재하다는 한계를 지닌다. 이에 본 연구는 기존의 다양한 논의들을 종합적으로 포괄하면서도 최근에 이뤄진 195편의 질적 연구(2000년-2016년의 기간에 Academy of Management Journal에 게재된 논문)의 방법론들을 아우를 수 있는 질적 연구 타당성 프레임워크를 제시하는 것을 목적으로 삼았다. 구체적으로 본 논문의 타당성 프레임워크는 연구 디자인 설계, 데이터 수집, 데이터 해석으로 이어지는 질적 연구 과정의 단계에서 고려해야 할 6가지 타당성 개념을 제안한다. 또한 본 논문은 이 프레임워크에서 제시하는 타당성 기준들을 79편의 질적연구논문(2012년-2016년의 기간에 Academy of Management Journal에 게재된 논문들)에 적용해봄으로써 그 실용성을 입증해보았다. 79편의 논문 분석 결과, 2012년-2016년 사이에 Academy of Management Journal에 게재된 질적 연구논문들은 그 타당성 수준이 해마다 감소하는 것으로 나타났다.

주요어: 타당성, 질적연구, 연구방법론