Management Accounting Research xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Management Accounting Research

journal homepage: www.elsevier.com/locate/mar

Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades

Maik Lachmann[a,*], Irene Trapp[b], Rouven Trapp[b]

[a] Technical University of Berlin, Department of Accounting and Management Control, Straße des 17. Juni 135, 10623 Berlin, Germany
[b] TU Dortmund University, Department of Accounting and Management Control, Vogelpothsweg 87, 44227 Dortmund, Germany

ARTICLE INFO

Article history:
Received 27 June 2015
Received in revised form 25 May 2016
Accepted 17 July 2016
Available online xxx

Keywords:
Method diversity
Positivist management accounting research
Theory diversity
Validity

ABSTRACT

This paper assesses the development and state of positivist management accounting research (PMAR). Based on a content analysis of 375 papers published in nine accounting journals over four decades, we conclude that a diverse set of research methods and theories, along with a consideration of validity, are necessary prerequisites for the accumulation of knowledge on management accounting (MA) practice. In light of diversity, we examine the studies with regard to their contents, methods and theoretical perspectives. Our analyses on validity comprise multiple facets of internal, external, construct and statistical conclusion validity. Regarding diversity, our findings suggest that PMAR has recently become narrower in terms of topics as it increasingly focuses on control issues. However, PMAR continues to rely on a variety of research methods and theoretical perspectives. Regarding validity, we find improvements for all four types of validity over time. However, potential for further progress persists. We discuss our findings in light of recent debates regarding the state of PMAR and highlight avenues for future research. Overall, we consider our study useful for assessing the discipline's achievements and evaluating its future paths.

1. Introduction

This paper analyses the diversity and validity in positivist management accounting research (PMAR) papers published in leading accounting journals over four decades. The study intends to advance our knowledge on the state and development of this field of research. We expect our paper to be useful for a critical assessment of past achievements and for a reflection on future avenues. In particular, recently expressed concerns regarding the path of management accounting (MA) research (e.g., Birnberg, 2009; Chow, 2010; Merchant, 2013; Salterio, 2015) emphasize the relevance of the insights provided by our study.

Our notion of diversity refers to the employment of different research methods and theoretical perspectives for investigations of MA practices (Birnberg et al., 1990). Employing different theoretical perspectives is important as each theory relies on specific assumptions and thus explains MA phenomena only partially (Hoque et al., 2013; Luft and Shields, 2002). Similarly, reliance on different methods is important due to the limitations that each method implies (Merchant and Van der Stede, 2006; Shields, 2015). Illuminating a MA phenomenon from different perspectives thus appears more likely to contribute to a comprehensive understanding rather than to remain entrenched in one (Chapman, 2012; Davilla and Oyon, 2008; Hopwood, 2008b; Merchant et al., 2003). Validity refers to the approximate truth of knowledge claims gathered through PMAR (Shadish et al., 2002, p. 34). Even if MA practices are illuminated from different perspectives but with considerable inherent flaws, diversity is unlikely to advance our understanding in the intended manner. Therefore, we consider diversity and validity primary dimensions for evaluating PMAR. Consequently, these dimensions serve as a backbone of the content analysis of the 375 PMAR papers presented in this study.

In light of recent debates on the state of MA research (e.g., Mittendorf, 2015; Scapens and Bromwich, 2010a), our evaluation focuses particularly on temporal trends. In so doing, we place the current state on a larger temporal scale to assess achievements and drawbacks. More precisely, we provide a comparative perspective on four periods. Our time frame reflects various watershed moments in the development of the discipline: The first period (1980–1982) captures the "empirical turn" of MA from a primarily normative discipline that relied on analytical modelling towards empirical research (Hopper et al., 2001; Klemstine and Maher, 1984). The second period (1990–1992) reflects the establishment of two academic MA journals propagating openness to diverse research approaches and thus potentially reinforcing diversity in

* Corresponding author.
E-mail addresses: maik.lachmann@tu-berlin.de (M. Lachmann), irene.trapp@tu-dortmund.de (I. Trapp), rouven.trapp@tu-dortmund.de (R. Trapp).

http://dx.doi.org/10.1016/j.mar.2016.07.002
1044-5005/

PMAR (Lindquist and Smith, 2009; Scapens and Bromwich, 2001). The third period (2000–2002) represents a time that is claimed to be characterized by diverse research approaches (Scapens, 2006). The fourth period (2010–2012) is expected to mirror an increasing topical and methodological narrowness (e.g., Birnberg, 2009; Hopwood, 2008a; Lukka, 2010; Merchant, 2013; Scapens and Bromwich, 2010b).

Our paper differs from previous studies primarily in two respects. First, we consider a plethora of research design choices beyond the classifications of topics, theories and methods. Thus, our study goes considerably further than previous bibliometric studies (e.g., Hesford et al., 2007; Lindquist and Smith, 2009; Scapens and Bromwich, 2010b). Moreover, it captures a longer time frame and a larger selection of journals. Second, our study complements a series of papers that discuss methodological issues (e.g., Ahrens and Chapman, 2006; Chenhall and Moers, 2007; Luft and Shields, 2014) by adopting an "ex post" perspective on the validity of published papers.[1] Our paper thus reflects the state of the art of PMAR and may raise awareness of issues that require consideration in future research (Brutus et al., 2013). Moreover, it provides an empirical contribution to the debate on whether PMAR is losing its openness to different research methods and relevance to practice (e.g., Chow, 2010; Krishnan, 2015; Merchant, 2010; Salterio, 2015; Tucker and Lowe, 2014).

Overall, our findings suggest that PMAR is becoming increasingly mature. Regarding diversity, our findings indicate that the range of topics studied by PMAR has become narrower over time. However, PMAR appears diversified, both in terms of research methods and theoretical perspectives. Thus, there is little evidence that PMAR follows the path of financial accounting research towards methodological and epistemological monism (e.g., Bonner et al., 2012; Oler et al., 2010). Regarding validity, our findings suggest that all types of validity tend to be increasingly addressed over time. However, there is still potential for substantial improvement. Correspondingly, we identify a series of topical as well as methodological issues that require further consideration.

This paper is structured as follows. In Section 2, we discuss diversity and validity and derive propositions on the development of PMAR based on the related literature. In Section 3, we explain our data collection and analysis. We present our findings in Section 4 and discuss them in Section 5.

2. Background and research propositions

2.1. Purpose and process of PMAR

Our study relies on the assumption that PMAR intends to acquire an in-depth understanding of MA practices (Malmi and Granlund, 2009; Van der Stede, 2015). PMAR focuses on causal explanations of MA phenomena that are common in many instances. Therefore, it draws inferences from a sample of specific observations to the general (Ittner, 2014; Luft and Shields, 2014).[2] Models of scientific enquiry (e.g., Gioia and Pitre, 1990; Kaplan, 1986; Snow and Thomas, 1994) suggest that the accumulation of knowledge typically begins with a description of a MA phenomenon and tends to move incrementally forward to its explanation (Kaplan, 1986). For this reason, research commonly develops measures that reflect the phenomenon of interest and explores its associations with other variables (Snow and Thomas, 1994). As research progresses, it culminates at best in the building of theories consisting of propositions that explain MA phenomena across a variety of conditions (Coloquitt and Zapata-Phelan, 2007; Malmi and Granlund, 2009). Commentaries on the development of MA research in the US and UK support this perspective. They suggest that PMAR was mostly descriptive after the "empirical turn" in the early 1980s and gradually became more explanatory (Maher, 2001; Otley, 2003; Scapens, 2006; Zimmerman, 2001). This understanding of the research process constitutes the background for the following propositions of how PMAR has evolved with regard to diversity and validity.[3]

2.2. Propositions on the development of diversity in PMAR

Knowledge building may benefit from the employment of different research methods to investigate a MA phenomenon because each is subject to strengths and limitations (e.g., Merchant and Van der Stede, 2006; Shields, 2015). We define the appropriation of different research methods as *method diversity* and expect that it increases in PMAR over time. First, the limitations of one method may serve as an initial point for future studies relying on other research designs. For instance, case studies allow for an in-depth exploration of MA innovations in their organizational settings (Davilla and Oyon, 2008; Merchant and Van der Stede, 2006). Although these studies may also contribute to theory refinement (Keating, 1995; Snow and Thomas, 1994), providing evidence on the wider applicability of propositions requires larger-scale methods (Birnberg et al., 1990; Modell, 2005). Survey studies, for example, may confirm associations among large samples of firms (Lillis and Mundy, 2005; Van der Stede et al., 2005). However, due to their cross-sectional design, most cannot establish causal inferences (Van der Stede, 2014). Employing different methods may offset the limitations of individual methods.

Second, we expect that the range of issues studied by PMAR, i.e., its *content diversity*, increases over time. If PMAR intends to develop an in-depth understanding of MA practice, we assume that developments in practice shape the research agendas (Baldvinsdottir et al., 2010; Mitchell, 2002). Therefore, the emergence of MA practices, such as the balanced scorecard (Kaplan, 1994), strategic MA (Langfield-Smith, 2008) or risk management and control (Soin and Collier, 2013; Van der Stede, 2011), may have broadened the agenda of PMAR. As the respective bodies of knowledge differ, we expect that different MA phenomena are investigated by employing diverse research methods. Therefore, we anticipate that method diversity increases over time. Indeed, anecdotal and bibliometric evidence suggests that PMAR diversified during the 1980s and 1990s regarding its research methods (Bhimani, 2002; Hesford et al., 2007; Hopper and Bui, 2016; Hopper et al., 2001; Scapens, 2006).

Knowledge building may also benefit from reliance on a variety of theoretical perspectives (e.g., Luft and Shields, 2002; Lukka and Granlund, 2002). For instance, Covaleski et al. (2003) compare budgeting research across the theoretical perspectives of economics, psychology, and sociology. They explain that each perspective implies distinctive research questions and refers to specific

[1] Exceptions include Bisbe et al. (2007), Hartmann and Moers (1999), Modell (2005) and Van der Stede et al. (2005), who review how particular methodological issues have been addressed by previously published papers.

[2] Our study excludes interpretive MA research because the types of validity considered in our study are typically only of concern to positivist researchers (Birnberg et al., 1990). By contrast, interpretive researchers focus on the "trustworthiness" of their research (Ahrens and Chapman, 2006; Davilla and Oyon, 2008).

[3] Following Kaplan (1986), we do not argue that the research process moves sequentially from one stage to another. In most cases, it will be iterative, as research findings at later stages may require theory refinement and further description of the MA phenomenon under study (Ferreira and Merchant, 1992). However, we expect that PMAR in general evolves as it accumulates established knowledge in particular areas over time. We expect that studies in these areas increasingly consist of explanatory research (see for these topics Krishnan, 2015; Shields, 2015; Van der Stede, 2015), whereas emerging issues are likely to be subject to more exploratory research (Merchant and Van der Stede, 2006).

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

levels of analysis. Reliance on one perspective thus provides only partial explanations for the phenomenon under study (Lukka and Mouritsen, 2002; Shields, 1997; Unerman and Chapman, 2014). For this reason, illuminating a MA phenomenon from different theoretical perspectives is likely to contribute to a more holistic understanding (Luft and Shields, 2002; Merchant et al., 2003). In this study, we refer to the employment of different theoretical perspectives as *theory diversity*. We argue that theory diversity in PMAR increases over time. Once knowledge is established about relationships informed by a particular theoretical perspective, additional insights may be generated based on illuminating the phenomenon from other perspectives. Theory diversity appears particularly likely to increase when empirical findings are inconsistent with previously employed perspectives. Commentaries on the development of MA research in Anglophone countries suggest that during its penetration in the early 1980s, PMAR was mostly based on economic theories (Hopper et al., 2001; Klemstine and Maher, 1984). This base then became more diversified as psychological and sociological theories were increasingly employed (Chenhall and Smith, 2011; Maher, 2001; Scapens, 2006). Correspondingly, bibliometric studies on MA research indicate an extensive reliance on economic, sociological and psychological perspectives in the 1980s and 1990s (Hesford et al., 2007; Hopper and Bui, 2016; Shields, 1997).

In summary, we expect that PMAR refers to an increasing range of topics and becomes more diversified with regard to methods and theoretical perspectives over time. However, our reliance on models of scientific enquiry disregards that a researcher's choice of topics and research approaches does not depend entirely on "curiosity" but also on personal incentives, i.e., tenure and promotion criteria (Gray and Milne, 2015; Hopwood, 2008a; Shields, 1997). This issue has been the subject of considerable debate in recent years (e.g., Birnberg, 2009; Chow, 2010; Hermanson, 2015; Lukka, 2010; Merchant, 2010). Career advancement in US business schools is said to be increasingly dependent on publishing in a small set of journals (Humphrey and Gendron, 2015; Merchant, 2013) and is progressively mimicked by European universities (Hopwood, 2008a; Messner, 2015; ter Bogt and Scapens, 2012). Papers published in these journals rely mainly on economic theory and draw on large archival datasets (Bonner et al., 2012; Merchant, 2010; Oler et al., 2010). Faced with the pressure to publish in these journals, a number of authors argue that researchers are incentivized to conduct research that best aligns – concerning both topics and research approaches – with papers previously published by the target journals (Chenhall and Smith, 2011; Maher, 2001; Merchant, 2010; Merchant, 2013; Hermanson, 2015). Consequently, content, method and theory diversity is likely to decline. Accordingly, such pressures put the aforementioned expectations into perspective. Taking our expectations based on models of scientific enquiry and the previously outlined concerns together, we specify the following propositions[4]:

Proposition 1. *Content diversity of PMAR increases among the 1980s, 1990s and 2000s and decreases between the 2000s and 2010s.*

Proposition 2. *Method diversity of PMAR increases among the 1980s, 1990s and 2000s and decreases between the 2000s and 2010s.*

Proposition 3. *Theory diversity of PMAR increases among the 1980s, 1990s and 2000s and decreases between the 2000s and 2010s.*

[4] For simplicity, we refer to "the 1980s" (etc.), which reflects the period 1980–1982. We acknowledge that this term does not imply a reference to the entire 1980–1989 period.

2.3. Propositions on the development of validity in PMAR

In addition to diversity, acquiring scientific knowledge about MA practice requires that PMAR investigates the intended phenomenon faithfully. In this context, *validity* refers to the "approximate truth" of the inferences from empirical investigations (Bisbe et al., 2007; Shadish et al., 2002, p. 35). The validity of inferences is a "core concern" for PMAR (Luft and Shields, 2014, p. 551) because it focuses on regularities among many instances but relies on a limited set of observations to draw inferences (Lukka and Kasanen, 1995). In the social sciences, evaluations of empirical research conventionally rely on four types of validity: internal validity, external validity, construct validity and statistical conclusion validity (Abernethy et al., 1999; Atkinson and Shaffir, 1998; Birnberg et al., 1990; Luft and Shields, 2014; Modell, 2005; Modell, 2009; Shields, 2015; Van der Stede et al., 2005).

Internal validity refers to whether the observed covariation between a dependent and independent variable reflects a causal relationship (Birnberg et al., 1990; Shadish et al., 2002, p. 53). Therefore, internal validity is a key issue for theory-testing studies that seek to explain MA phenomena (Snow and Thomas, 1994; Van der Stede et al., 2005). If PMAR gradually moves from describing to explaining MA practices, as argued in Section 2.1, increasing emphasis should be placed on internal validity. Consequently, we expect that the internal validity of PMAR increases over time. *External validity* refers to whether causal relations identified in one study can be generalized to wider settings and populations (Birnberg et al., 1990; Shadish et al., 2002, p. 83). Models of scientific enquiry suggest that explanatory research first investigates correlations among variables under specific conditions and then moves on by building theories with more general propositions and wider applicability (Kaplan, 1986). For this reason, we argue that the external validity of PMAR increases over time as research progressively aims at testing theories that are applicable to a variety of settings.

Construct validity refers to whether an operational definition of a construct is an adequate measure of the underlying theoretical concept (Abernethy et al., 1999; Shadish et al., 2002, p. p. 65). Constructs play a key role in theory building and testing (Hamann et al., 2013; Shadish et al., 2002, p. 100). Therefore, we expect that researchers devote increasing attention to construct validity as PMAR becomes more explanatory. A progressing complexity of the relations investigated reinforces this assumption (Luft and Shields, 2003). Whereas early explanatory studies regularly focus on core propositions, subsequent research considers additional variables, thus resulting in more complex empirical models (Colquitt and Zapata-Phelan, 2007; Snow and Thomas, 1994). *Statistical conclusion validity* refers to the use of appropriate statistics for inferences about the covariation between dependent and independent variables (Shadish et al., 2002, p. 37). Associations between variables may be poorly estimated in the presence of threats such as endogeneity or multicollinearity (Ittner, 2014; Nikolaev and van Lent, 2005). In particular, investigating more complex relationships is likely to give rise to such issues (Chenhall and Moers, 2007). Therefore, we expect that researchers increasingly test for these effects to avoid biased estimations of associations, implying that statistical conclusion validity increases over time.

In summary, we argue that the assumed tendency of PMAR to move from primarily descriptive to more explanatory research affects the four dimensions of validity that play pivotal roles in building and testing generally applicable propositions. In contrast to the diversity dimensions, the literature does not include arguments that put our expectations into perspective. For this reason, we specify the following propositions:

Proposition 4. *The internal validity of PMAR increases over time periods.*

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

Proposition 5. *The external validity of PMAR increases over time periods.*

Proposition 6. *The construct validity of PMAR increases over time periods.*

Proposition 7. *The statistical conclusion validity of PMAR increases over time periods.*

3. Method

3.1. Selection of journals and papers

Our journal selection follows that of Hesford et al. (2007), who focus on "outlets in which management accounting research has been prominently published" (p. 5). This selection also reflects the leading journals according to accounting faculty surveys and journal rankings (e.g., Ballas and Theoharakis, 2003; Bonner et al., 2006). However, we modified the selection in three respects. Since we focus on positivist research, we excluded the Journal of Accounting Literature due to its previously exclusive focus on literature reviews. For the same reason, we excluded the Review of Accounting Studies, which primarily publishes MA research that draws on mathematical modelling (Hesford et al., 2007). We enhanced the journal selection by adding the European Accounting Review because of its significant reputation as the leading journal of the European Accounting Association and its important role in disseminating a diverse set of MA research papers (Bhimani, 2002; Raffournier and Schatt, 2010; Van Campenhout and Van Caneghem, 2010). Thus, our final selection comprises the following nine journals (in alphabetical order): Accounting, Organizations and Society (AOS), Behavioral Research in Accounting (BRIA), Contemporary Accounting Research (CAR), European Accounting Review (EAR), Journal of Accounting and Economics (JAE), Journal of Accounting Research (JAR), Journal of Management Accounting Research (JMAR), Management Accounting Research (MAR) and The Accounting Review (TAR). For the reasons outlined in Section 1, we selected papers published in four periods, each consisting of three years (1980–1982, 1990–1992, 2000–2002, 2010–2012).[5]

Due to their exclusive focus on MA, we included all full papers published in JMAR and MAR during the aforementioned periods that employ a positivist research approach.[6] Concerning the remaining seven general accounting journals, two authors independently read the abstracts of all full papers to identify PMAR papers. In this context, we relied on the definition of MA proposed by Foster and Young (1997).[7] This approach resulted in a high degree of consensus. In case of disagreement, all three authors reviewed the paper and made a joint decision. The selection processes identified 375 PMAR papers. These papers are almost evenly divided between North American (52.8%) and European journals (47.2%).[8]

3.2. Content analysis

The papers were subject to a content analysis. The corresponding category scheme primarily draws on the framework by Scandura and Williams (2000), who apply it in a comprehensive assessment of organizational studies. We refer to this framework as it comprises various dimensions that reflect our research interest in diversity and validity. Table 1 provides an overview of the corresponding coding dimensions. Each coding dimension consists of categories to which we assigned the studies. Whereas the selection of coding dimensions is derived from Scandura and Williams (2000), we developed the respective category schemes based on the related MA literature (e.g., Hesford et al., 2007; Lillis, 1999; Shields, 1997; Van der Stede et al., 2005). The rationales for the coding dimensions and the description of the corresponding categories are disclosed in full detail in Appendix A.

Two of the authors performed the coding procedures independently based on coding guidelines that included explanations of the categories and examples. First, both authors conducted a precoding based on a randomized sample of 40 papers, compared their codings and discussed deviations. This review led to few clarifications regarding the category definitions. Subsequent to these modifications, the authors performed the final coding. Codings were reviewed and compared after 50 papers were completed. Any deviation was subject to discussion until the authors reached an agreement that was consistent with the treatment of the other papers.

3.3. Data analysis

In line with Scandura and Williams (2000), our data analysis relies on a series of descriptive statistics and statistical tests. For nominal data, we performed chi-square tests to identify associations between particular coding dimensions. We performed multi-sample median tests and Kruskal-Wallis tests to determine whether differences among ordinal data from the four periods were significant. In addition, we used linear regression analyses to determine whether changes among periods reflect continuity (Scandura and Williams, 2000).

The degree of content, method and theory diversity was measured by calculating "heterogeneity indices" (Harrison and Klein, 2007; Scandura and Williams, 2000) that reflect the range of topics, methods and theoretical perspectives. These indices complement our descriptive analyses. We calculated these indices using the following formula (Blau, 1977; p. 78):

$$h_{j,k} = 1 - \sum_{i=1}^{n} p_i^2$$

where $h_{j,k}$ represents the heterogeneity index for code j in period k. p_i is the fraction of the content-analysed studies assigned to category i of code j with n as the number of categories in code j. The values of the index range from zero (minimum heterogeneity) to (n – 1)/n (maximum heterogeneity). The latter implies that the studies are evenly divided among the categories of a particular code

[5] We concede that the non-random sampling does not allow for inferences regarding other years. Instead, our study provides a comparison of four distinctive periods. We also acknowledge that the periods may synchronize with the tenures of individual editors. Therefore, differences among the periods may be driven by varying editorial preferences. For this reason, we identified the editors in charge of the respective journals for each year and checked whether rotations occurred. For most journals with limited editor tenure, rotations occurred during the individual periods. Therefore, we expect this issue to have limited impact on our findings.

[6] The research methods considered correspond with the categories distinguished for the content analysis as explained in Appendix A (Code 2).

[7] According to Foster and Young (1997, p. 64), MA constitutes a "value adding, continuous improvement process of planning, designing, measuring, and operating nonfinancial and financial information systems that guides management action, motivates behaviour, and supports and creates the cultural values necessary to achieve an organization's strategic, tactical and operative objectives".

[8] Our journal selection facilitates comparisons between European and North American journals, which may address claims that the latter contribute considerably to an increasing narrowness of MA research (e.g., Merchant, 2010). As supplementary analyses, we therefore performed each analysis separately for European and North American journals. If considerable differences emerged, we report them in Section 4.

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

Table 1
Content analysis coding dimensions derived from the Scandura and Williams (2000) framework.

Dimension of analysis	Type of diversity/validity	Coding dimension (Code in brackets)	Description[a]
Diversity	Content diversity	Content (1)	Refers to the subject areas studied in the papers.
	Method diversity	Research method (2)	Refers to the methods employed in the studies.
	Theory diversity	Theoretical perspective (3)	Refers to the theoretical perspectives employed in the studies.
Supplementary dimension of analysis		Level of analysis (4)	Refers to the level at which the issue under study occurs.
Validity	Internal validity	Time frame (5)	Refers to whether data were collected cross-sectionally or longitudinally.
	External validity	Type of sample (6)	Refers to the setting of the studies.
		Primary occupation (7)	Refers to the professional background of study participants.
	Construct validity	Number of measures for construct validation (8)	Refers to the measures taken to address construct validity as disclosed in the studies.
		Number of reliability measures (9)	Refers to the measures taken to address reliability of research designs as disclosed in the studies.
		Type of dependent variables (10)	Refers to the types of dependent variables employed in the studies.
		Number of data sources (11)	Refers to the number of data sources used in the studies.
	Statistical conclusion validity	Consideration of threats to statistical conclusion validity (12)	Refers to the measures taken to address threats that may lead to incorrect statistical conclusions as disclosed in the studies.

[a] Explanations of coding dimensions and corresponding categories are disclosed in full detail in Appendix A.

(Harrison and Klein, 2007). We draw primarily on these indices for time comparisons. An increasing heterogeneity index suggests increasing diversity, which may enhance our understanding of MA practices.

4. Findings

In the remainder of this section, we first address the issue of diversity and present the results for Propositions 1–3 that refer to content, method and theory diversity (Section 4.1). In the second step, we focus on validity and provide findings for the Propositions 4–7 (Section 4.2). In the sub-sections that refer to the respective propositions, we provide a series of analyses as previously outlined.

4.1. Findings on the development of diversity in PMAR

4.1.1. Content diversity (Proposition 1)

Proposition 1 refers to the range of topics studied and suggests an increase in content diversity among the first three decades and a decrease between the 2000s and 2010s. Table 2 reports the classification of papers according to their content (Code 1). In this context, we rely on a slightly modified category scheme reported by Hesford et al. (2007). In the 1980s, PMAR investigates almost evenly decision-facilitating (category "Cost and planning"; 45.8% of the papers) as well as decision-influencing (category "Control"; 54.2%) issues. In the two following periods, the shares of papers from both categories decrease, whereas the share of papers on other topics grows. However, these tendencies reverse in the 2010s as control issues are increasingly addressed, with a particular emphasis placed on performance measurement (43.8% of all papers published in this period). Following Scandura and Williams (2000), we conducted time-based regressions for all coding dimensions to detect whether changes over time represent a linear pattern that indicates some continuity in the developments. For this reason, we performed simple linear regression analyses for each category. The year of publication is the dependent variable and a dummy for the respective sub-category is the independent variable.[9] The regression results underline the aforementioned development as we find that the decrease in cost and planning studies and the increase in control studies are significant at the 1% and 5% levels, respectively. In addition, we detect significant developments for various sub-categories (e.g., budgeting or performance measurement and evaluation), as shown in Table 2 (column "Overall trend"). Correspondingly, the heterogeneity index shown in Table 2 increases during the first three periods (1980–1982: 0.75; 1990–1992: 0.83; 2000–2002: 0.86) and declines in the last period (2010–2012: 0.76). Overall, these findings support Proposition 1.[10]

[9] For instance, the regression for the category "Control" employs a dummy variable as the independent variable that equals 1 for papers assigned to any sub-category of "Control" and 0 otherwise. Increases or decreases refer to the number of studies. We conducted equivalent analyses for each of the codes except for codes 8 and 9. We report significant findings in the tables (column "Overall trend"). If we do not report a tendency for a category, findings were insignificant.

[10] Supplementary analyses of papers published in North American and European journals reveal that the overall tendency is reflected by both journal groups. Moreover, the European journals (average heterogeneity index (AHI): 0.58) are not considerably more diversified than the North American ones (AHI: 0.55).

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

Table 2
Contents.

Content (Code 1)[†]	1980–1982 # (%)	1990–1992 # (%)	2000–2002 # (%)	2010–2012 # (%)	Total	Over-all trend
Cost and planning						
Cost allocation	0[a] (0.0%)[b]	3 (3.8%)	12 (10.7%)	12 (7.5%)	27	
Other cost accounting topics	2 (8.3%)	6 (7.6%)	6 (5.4%)	1 (0.6%)	15	*** (−)
Cost practices	0 (0.0%)	4 (5.1%)	3 (2.7%)	10 (6.3%)	17	
Budgeting	7 (29.2%)	12 (15.2%)	13 (11.6%)	8 (5.0%)	40	*** (−)
Capital budgeting	2 (8.3%)	2 (2.5%)	2 (1.8%)	6 (3.8%)	12	
Total "Cost and planning"	**11 (45.8%)**	**27 (34.2%)**	**36 (32.1%)**	**37 (23.1%)**	**111**	**** (−)**
Control						
Performance measurement and evaluation	8 (33.3%)	27 (34.2%)	27 (24.1%)	70 (43.8%)	132	* (+)
Organizational control	5 (20.8%)	4 (5.1%)	17 (15.2%)	25 (15.6%)	51	
International control	0 (0.0%)	5 (6.3%)	6 (5.4%)	0 (0.0%)	11	* (−)
Multiple	0 (0.0%)	2 (2.5%)	2 (1.8%)	6 (3.8%)	10	
Total "Control"	**13 (54.2%)**	**38 (48.1%)**	**52 (46.4%)**	**101 (63.1%)**	**204**	**** (+)**
Other						
Quality (TQM)	0 (0.0%)	0 (0.0%)	1 (0.9%)	0 (0.0%)	1	
Just-in-time	0 (0.0%)	1 (1.3%)	2 (1.8%)	0 (0.0%)	3	
Strategic management	0 (0.0%)	9 (11.4%)	16 (14.3%)	18 (11.3%)	43	
Transfer Pricing	0 (0.0%)	2 (2.5%)	3 (2.7%)	0 (0.0%)	5	
Multiple	0 (0.0%)	2 (2.5%)	2 (1.8%)	4 (2.5%)	8	
Total "Other"	**0 (0.0%)**	**14 (17.7%)**	**24 (21.4%)**	**22 (13.8%)**	**60**	
Heterogeneity indices	0.75	0.83	0.86	0.76		

Refers to the number of papers (n = 375). [†] Sub-categories (6), (11) and (14) are not tabulated since no papers have been assigned to them. [a] Absolute number of papers. [b] Percentage of papers from the respective category out of the total number of papers published within the period. Findings from time-based regression analyses: ***—significant decrease (−) ($p < 0.01$); **—significant increase (+) ($p < 0.05$); *—significant increase (+) or decrease (−) ($p < 0.1$).

Table 3
Research methods.

Panel A: Research method (Code 2) per period

	1980–1982 # (%)	1990–1992 # (%)	2000–2002 # (%)	2010–2012 # (%)	Total	Over-all trend
Archival	0[a] (0.0%)[b]	13 (15.9%)	24 (19.5%)	46 (26.9%)	83	**(+)
Survey	6 (24.0%)	27 (32.9%)	39 (31.7%)	38 (22.2%)	110	
Experimental[c]	15 (60.0%)	23 (28.0%)	27 (22.0%)	50 (29.2%)	115	*(−)
Case study	2 (8.0%)	8 (9.8%)	10 (8.1%)	18 (10.5%)	38	
Field study	2 (8.0%)	11 (13.4%)	23 (18.7%)	19 (11.1%)	55	
Heterogeneity indices	0.57	0.76	0.77	0.77		

Panel B: Research method (Code 2) per content group (Code 1)

	Cost and planning # (%)	Control # (%)	Other# (%)	Total
Archival	27[d] (23.5%)[e]	50 (22.8%)	6 (9.0%)	83
Survey	39 (33.9%)	56 (25.6%)	15 (22.4%)	110
Experimental	30 (26.1%)	65 (29.7%)	20 (29.9%)	115
Case study	10 (8.7%)	16 (7.3%)	12 (17.9%)	38
Field study	9 (7.8%)	32 (14.6%)	14 (20.9%)	55
Heterogeneity indices	0.75	0.77	0.77	

Refers to the number of samples (n = 401). *Panel A:* [a] Absolute number of samples. [b] Percentage of samples from the respective category out of the total number of samples published within the period. Findings from time-based regression analyses: ***—significant increase (+) ($p < 0.01$); *—significant decrease (−) ($p < 0.1$). *Panel B:* [c] Includes two field experiments. [d] Absolute number of samples employing the respective research method per content group. [e] Percentage of samples from the respective research method out of the total number of samples assigned to the respective content group. Chi-square test for cross-tabulation of methods and content groups: $\chi^2 = 20.158$; $p = 0.01$; n = 401.

4.1.2. Method diversity (Proposition 2)

According to Proposition 2, method diversity increases over the first three time periods but declines in the fourth period. In reference to the categories of Hesford et al. (2007), Table 3 (Panel A) shows the spreads of research methods (Code 2) in the respective periods. The table reveals that experimental studies represent 60.0% of PMAR in the 1980s, which corresponds with a comparatively low heterogeneity index of 0.57. However, since the 1990s, method diversity has grown substantially due to an increasing reliance on archival, survey, and field studies. From the 1990s onwards, we find a balanced spectrum of research approaches, which is also reflected by high and stable heterogeneity indices (1990–1992: 0.76; 2000–2002 and 2010–2012: 0.77). Time-based regression analyses indicate a significant increase in the reliance on archival studies ($p < 0.01$) and a significant decrease in experimental research ($p < 0.1$).

In Section 2.2, we argued that method diversity implies the potential to overcome the limitations of other research approaches. Therefore, it seems important that not only PMAR in general but also specific subject areas are diversified. For this reason, we cross-tabulate the subject areas (i.e., cost and planning, control and other) and the research methods employed. In this way, we illuminate whether the overall method diversity is mirrored in the main subject areas. The findings reported in Table 3 (Panel B) indicate a comparatively strong reliance on archival, survey and experimental research for cost and planning as well as control issues. In contrast, research on other topics tends to rely more strongly on case and field studies. Nonetheless, the heterogeneity indices per subject area (ranging from 0.75 to 0.77) suggest a considerable diversity in all of them. In addition, untabulated findings indicate that method diversity increases over time or is at least stable in all three subject

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

areas. Therefore, we conclude that the main subject areas show a high method diversity.[11]

In summary, our findings corroborate Proposition 2 only partially. We find evidence for an increasing method diversity from the 1980s to the 1990s and almost constant method diversity in the following time periods. Accordingly, our findings do not indicate the anticipated decline in method diversity from the 2000s to the 2010s.

4.1.3. Theory diversity (Proposition 3)

Proposition 3 suggests that theory diversity will follow a similar path as method diversity. Table 4 summarises our classification of studies with regard to their theoretical perspectives (Code 3) based on the category scheme reported by Shields (1997). Panel A shows that studies published in the 1980s draw mostly on psychological theories to explain their findings, whereas a plethora of theoretical perspectives is used in the following decades. More precisely, over the last three decades PMAR most frequently draws on economic (between 15.9% and 22.8%) and psychological theories (between 15.4% and 22.8%) and theories of organizational behaviour (between 23.2% and 27.5%). Time-based regressions reveal significant increases regarding organizational behaviour ($p < 0.05$) and production and operations management ($p < 0.1$), as well as a significant decrease in studies without an explicit theoretical perspective ($p < 0.01$). The high and stable heterogeneity indices from the 1990s onwards (1990–1992: 0.82; 2000–2002: 0.81; 2010–2012: 0.79) indicate strong and persistent theory diversity. In other words, they do not indicate an increasing narrowness in terms of theoretical perspectives.

In Section 2.2, we argued that the reliance on different theoretical perspectives may reinforce our understanding of MA practice. We thus analyse whether the main subject areas show a similar degree of theory diversity as PMAR in general. For this reason, we cross-tabulate the theoretical perspectives and the subject areas as shown in Table 4 (Panel B). The heterogeneity indices per subject area differ only slightly (ranging from 0.78 to 0.83) and suggest that multiple theoretical perspectives are used in each of these subject areas. Nevertheless, the analysis reveals area-specific patterns, such as the infrequent reference of studies on other topics to economic theories or the low share of control topics that do not rely on an explicit theoretical perspective. Untabulated time comparisons show that theory diversity increases in the areas of cost and planning and other topics, whereas it slightly declines in the control area since the 1990s.[12] Similarly to the case of method diversity, we find partial support for Proposition 3. From the 1990s onwards, theory diversity is high and largely stable. Hence, we do not find evidence for a substantial decrease in theory diversity in the 2010s.

4.1.4. Supplementary analysis on levels of analysis

In Section 2.2, we argued that different theoretical perspectives focus on different levels of analysis (Covaleski et al., 2003; Luft and Shields, 2003). For a further assessment of PMAR, we illuminate whether the developments concerning the theoretical perspectives imply changes in the levels of analysis. For this reason, we follow Scandura and Williams (2000) and supplement our analysis on diversity with an investigation of these levels. In line with Luft and Shields (2003), we refer to the dependent and independent variables investigated and define the level of analysis (Code 4) as the one "at which the variation of interest occurs" (p. 175). If no variables were employed, we identified the level of analysis based on the respective research questions. If at least two different levels were addressed, we assigned the study to the cross-level category.

Table 5 (Panel A) reports developments over time and shows that PMAR primarily focuses on the individual level in the 1980s. In the subsequent decades, it has shifted more towards the organizational and the beyond-organizational levels. Most studies refer to one level of analysis, whereas cross-level studies account for roughly 10% of the studies. Time-based regressions reveal significant increases for the organizational and beyond-organizational levels ($p < 0.1$ and $p < 0.05$ respectively) and a significant decrease in studies at the individual level ($p < 0.01$).

Our illumination of the levels of analysis per subject area (Table 5, Panel B) reveals two notable patterns. First, PMAR on control issues focuses more strongly on the individual level than the other two subject areas. Second, we find that over time, PMAR on cost and planning issues refers increasingly to the beyond-organizational level. Overall, this supplementary analysis suggests a gradual shift in the focus of PMAR from the individual to the organizational level and beyond.

4.2. Findings on the development of validity in PMAR

4.2.1. Internal validity (Proposition 4)

Proposition 4 suggests that internal validity increases over time. Following Scandura and Williams (2000), we address internal validity by examining the time frames (Code 5) of the studies. Longitudinal studies tend to imply a higher level of internal validity than cross-sectional ones because they allow researchers to directly identify causal relationships in which causes precede effects (Van der Stede, 2014). Table 6 (Panel A) indicates that the share of longitudinal studies decreases among the first three periods and increases from the 2000s to the 2010s. This finding thus corresponds only partially with Proposition 4.

In a second step, we illuminate whether this development is driven by particular research methods. Case and field studies (Ahrens and Dent, 1998), archival (Moers, 2007) and experimental research are frequently longitudinal (Maines et al., 2006).[13] By contrast, Van der Stede et al. (2005) find that MA surveys are usually cross-sectional, although most "aim to test theories that specify causal relationships among variables" (p. 665). For this reason, cross-tabulating time frames and research methods may shed light on the question regarding whether PMAR has overcome the latter shortcoming. The findings presented in Table 6 (Panel B) confirm that archival, experimental and case studies mostly gather longitudinal data. They account for 88.2% of the longitudinal studies, whereas surveys that collect data at two or more points in time remain a rare exception over all time periods. In summary, we find limited support for Proposition 4. Moreover, we conclude that a considerable limitation with regard to survey research prevails.

4.2.2. External validity (Proposition 5)

According to Proposition 5, external validity increases over time. In line with Scandura and Williams (2000), we focus on the

[11] Untabulated comparisons between North American and European accounting journals reveal that the research methods most frequently applied in studies published in North American journals are archival and experimental approaches, whereas the methods most frequently published in European journals are surveys, case and field studies. However, contrary to the views expressed in the literature (e.g., Callen, 2015; Merchant, 2010), European journals (AHI: 0.64) do not appear to be considerably more diversified than North American journals (AHI: 0.61).

[12] A supplementary comparison of North American and European journals shows that studies published in North American (European) journals rely mainly on economic and psychological (sociological and organizational behaviour) theories. In both groups, theory diversity increases from the 1980s to 1990s and slightly decreases thereafter. Moreover, theory diversity appears slightly higher in North American journals (AHI: 0.76) than in European journals (AHI: 0.72).

[13] Note that we consider experimental studies longitudinal even if data were collected at one point in time because the manipulation of independent variables precedes the observation of the dependent variables. Therefore, causality can be established.

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

Table 4
Theoretical perspectives.

Panel A: Theoretical perspective (Code 3) per period

	1980–1982 # (%)	1990–1992 # (%)	2000–2002 # (%)	2010–2012 # (%)	Total	Over-all trend
Economics	4[a] (16.0%)[b]	13 (15.9%)	20 (16.3%)	39 (22.8%)	76	
Psychology	12 (48.0%)	16 (19.5%)	19 (15.4%)	39 (22.8%)	86	
Sociology	2 (8.0%)	9 (11.0%)	18 (14.6%)	24 (14.0%)	53	
Organizational behaviour	0 (0.0%)	19 (23.2%)	36 (29.3%)	47 (27.5%)	102	**(+)
Production and operations management	0 (0.0%)	8 (9.8%)	4 (3.3%)	3 (1.8%)	15	*(+)
Multiple	2 (8.0%)	1 (1.2%)	3 (2.4%)	4 (2.3%)	10	
No theory	5 (20.0%)	16 (19.5%)	23 (18.7%)	15 (8.8%)	59	***(−)
Heterogeneity indices	0.69	0.82	0.81	0.79		

Panel B: Theoretical perspective (Code 3) per content group (Code 1)

	Cost and planning # (%)	Control # (%)	Other # (%)	Total
Economics	20[c] (17.4%)[d]	53 (24.2%)	3 (4.5%)	76
Psychology	25 (21.7%)	51 (23.3%)	10 (14.9%)	86
Sociology	17 (14.8%)	29 (13.2%)	7 (10.4%)	53
Organizational behaviour	19 (16.5%)	64 (29.2%)	19 (28.4%)	102
Production and operations management	7 (6.1%)	2 (0.9%)	6 (9.0%)	15
Multiple	3 (2.6%)	5 (2.3%)	2 (3.0%)	10
No theory	24 (20.9%)	15 (6.8%)	20 (29.9%)	59
Heterogeneity indices	0.83	0.78	0.79	

Refers to the number of samples (n = 401). *Panel A:* [a] Absolute number of samples. [b] Percentage of samples from the respective category out of the total number of samples published within the period. Findings from time-based regression analyses: ***—significant decrease (−) ($p < 0.01$); **—significant increase (+) ($p < 0.05$); *—significant increase (+) ($p < 0.1$). *Panel B:* [c] Absolute number of samples relying on the respective theoretical perspective per content group. [d] Percentage of samples from the respective category out of the total number of samples assigned to the respective content group. Chi-square test for cross-tabulation of theoretical perspectives and content groups: $\chi^2 = 52.027$; $p < 0.001$; n = 401.

Table 5
Levels of analysis.

Panel A: Level of analysis (Code 4) per period

	1980–1982 # (%)	1990–1992 # (%)	2000–2002 # (%)	2010–2012 # (%)	Total	Over-all trend
Individual	21[a] (84.0%)[b]	30 (36.6%)	33 (26.8%)	58 (33.9%)	142	***(−)
Subunit	1 (4.0%)	8 (9.8%)	12 (9.8%)	9 (5.3%)	30	
Organization	2 (8.0%)	31 (37.8%)	59 (48.0%)	67 (39.2%)	159	*(+)
Beyond organization	0 (0.0%)	3 (3.7%)	8 (6.5%)	17 (9.9%)	28	**(+)
Cross-level	1 (4.0%)	10 (12.2%)	11 (8.9%)	20 (11.7%)	42	

Panel B: Level of analysis (Code 4) per content group (Code 1)

	Cost and planning # (%)	Control # (%)	Other # (%)	Total
Individual	34[c] (29.6%)[d]	92 (42.0%)	16 (23.9%)	142
Subunit	14 (12.2%)	9 (4.1%)	7 (10.4%)	30
Organization	46 (40.0%)	80 (36.5%)	33 (49.3%)	159
Beyond organization	11 (9.6%)	15 (6.8%)	2 (3.0%)	28
Cross-level	10 (8.7%)	23 (10.5%)	9 (13.4%)	42

Refers to the number of samples (n = 401). *Panel A:* [a] Absolute number of samples. [b] Percentage of samples from the respective category out of the total number of samples published within the period. Findings from time-based regression analyses: ***—significant decrease (−) ($p < 0.01$); **—significant increase (+) ($p < 0.05$); *—significant increase (+) ($p < 0.1$). *Panel B:* [c] Absolute number of samples referring to the respective level of analysis per content group. [d] Percentage of samples from the respective category out of the total number of samples assigned to the respective content group. Chi-square test for cross-tabulation of levels of analysis and content groups: $\chi^2 = 19.463$; $p = 0.01$; n = 401.

types of samples and the primary occupation of participants as coding dimensions that shed light on the potential to generalize findings (Dahlstrom et al., 2008). Table 7 reports our findings on the types of sample (Code 6) based on the category scheme by Scapens and Bromwich (2010b). "Generic" refers to laboratory experiments, whereas the other categories represent the remaining research methods. The findings suggest that the share of studies on manufacturing decreases in favour of studies that focus on either specific industries, the service sector or both (subsumed under other types of samples). These tendencies are emphasized by time-based regressions that reveal significant increases for studies in the public sector or other settings ($p < 0.05$ and $p < 0.01$ respectively) and significant decreases for studies in a generic or manufacturing setting ($p < 0.1$ and $p < 0.01$ respectively). Overall, PMAR seems to refer to an increasing variety of settings, which may facilitate the identification of differences and commonalities and increase the generalizability of findings among different sectors.

Because external validity also refers to generalizing findings among different individuals, groups and occupations (Brutus et al., 2013; Scandura and Williams, 2000), we capture the primary occupation (Code 7) of the participants. For this reason, we refer to the categories reported by Scandura and Williams (2000), which we adapted to the MA context. Note that this analysis excludes archival studies that do not involve participants. Table 8 (Panel A) reports the findings on primary occupations per period. We detect several notable tendencies. First, the share of studies involving management accountants and accountants is low (maximal 8.0%) over the entire period and tends to decline. Second, we identify an increasing number of studies with top-level managers as participants. Third, we detect an increase in the number of studies that involve multiple professional groups (up to 29.6% in the 2010s). Moreover, time-

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

Table 6
Time frames.

Panel A: Time frame (Code 5) per period

	1980–1982# (%)	1990–1992# (%)	2000–2002# (%)	2010–2012# (%)	Total
Cross-sectional	7[a] (28.0%)[b]	36 (43.9%)	60 (48.8%)	60 (35.1%)	163
Longitudinal	18 (72.0%)	46 (56.1%)	63 (51.2%)	111 (64.9%)	238

Panel B: Time frame (Code 5) per research method (Code 2)

	Cross-sectional# (%)	Longitudinal# (%)	Total
Archival	18[c] (21.7%)[d]	65 (78.3%)	83
Survey	104 (94.5%)	6 (5.5%)	110
Experimental	0 (0.0%)	115 (100.0%)	115
Case study	8 (21.1%)	30 (78.9%)	38
Field study	33 (60.0%)	22 (40.0%)	55

Refers to the number of samples (n = 401). *Panel A:* [a] Absolute number of samples. [b] Percentage of samples from the respective category out of the total number of samples published within the period. *Panel B:* [c] Absolute number of samples with the respective time frame per research method. [d] Percentage of samples with the respective time frame out of the total number of samples assigned to the respective research method. Chi-square test for cross-tabulation of research methods and time frames: $\chi^2 = 238.164$; $p < 0.001$; n = 401.

Table 7
Types of sample.

Type of sample (Code 6) per period

	1980–1982 # (%)	1990–1992 # (%)	2000–2002 # (%)	2010–2012 # (%)	Total	Over-all trend
Generic[c]	15[a] (60.0%)[b]	23 (28.0%)	26 (21.1%)	49 (28.7%)	113	*(−)
Manufacturing	5 (20.0%)	19 (23.2%)	13 (10.6%)	10 (5.8%)	47	***(−)
Specific industries	1 (4.0%)	16 (19.5%)	25 (20.3%)	15 (8.8%)	57	
Services	2 (8.0%)	3 (3.7%)	5 (4.1%)	17 (9.9%)	27	
Public sector	2 (8.0%)	3 (3.7%)	13 (10.6%)	23 (13.5%)	41	**(+)
Not-for-profit	0 (0.0%)	4 (4.9%)	4 (3.3%)	5 (2.9%)	13	
Specific countries	0 (0.0%)	6 (7.3%)	7 (5.7%)	5 (2.9%)	18	
Multinationals	0 (0.0%)	1 (1.2%)	4 (3.3%)	6 (3.5%)	11	
Other[d]	0 (0.0%)	7 (8.5%)	26 (21.1%)	41 (24.0%)	74	***(+)

Refers to the number of samples (n = 401). [a] Absolute number of samples. [b] Percentage of samples from the respective category out of the total number of samples published within the period. [c] Comprises laboratory experiments conducted in an abstract setting and does not equal with the number of experimental studies reported in Table 2 as the latter category includes two field experiments that are assigned to non-generic types of samples. [d] Primarily refers to samples that comprise manufacturing as well as service firms. Findings from time-based regression analyses: ***—significant increase (+) or decrease (−) ($p < 0.01$); **—significant increase (+) ($p < 0.05$); *—significant decrease (−) ($p < 0.1$).

Table 8
Primary occupations.

Panel A: Primary occupation (Code 7) per period

	1980–1982 # (%)	1990–1992 # (%)	2000–2002 # (%)	2010–2012 # (%)	Total	Over-all trend
Management accountants and accountants	2[a] (8.0%)[b]	4 (5.8%)	5 (5.1%)	4 (3.2%)	15	
Top-level management	1 (4.0%)	4 (5.8%)	9 (9.1%)	16 (12.8%)	30	
Other management	7 (28.0%)	22 (31.9%)	16 (16.2%)	20 (16.0%)	65	***(−)
Others	1 (4.0%)	3 (4.3%)	7 (7.1%)	4 (3.2%)	15	
Multiple	7 (28.0%)	13 (18.8%)	27 (27.3%)	37 (29.6%)	84	
Undergraduate students	2 (8.0%)	11 (15.9%)	14 (14.1%)	18 (14.4%)	45	
Graduate students/graduates	4 (16.0%)	3 (4.3%)	8 (8.1%)	23 (18.4%)	38	
No report	1 (4.0%)	9 (13.0%)	13 (13.1%)	3 (2.4%)	26	**(−)

Panel B: Primary occupation (Code 7) per research method

	Survey	Experimental	Case study	Field Study	Total
Management accountants and accountants	14[c] (12.7%)[d]	0 (0.0%)	0 (0.0%)	1 (1.8%)	15
Top-level manangement	15 (13.6%)	2 (1.7%)	3 (7.9%)	10 (18.2%)	30
Other management	41 (37.3%)	4 (3.5%)	10 (26.3%)	10 (18.2%)	65
Others	8 (7.3%)	4 (3.5%)	0 (0.0%)	3 (5.5%)	15
Multiple	19 (17.3%)	21 (18.3%)	19 (50.0%)	25 (45.5%)	84
Undergraduates students	0 (0.0%)	45 (39.1%)	0 (0.0%)	0 (0.0%)	45
Graduate students/graduates	0 (0.0%)	38 (33.0%)	0 (0.0%)	0 (0.0%)	38
No report	13 (11.8%)	1 (0.9%)	6 (15.8%)	6 (10.9%)	26

Refers to the number of samples that involve participants and do not rely exclusively on archival data (n = 318; archival studies are excluded). *Panel A:* [a] Absolute number of samples. [b] Percentage of samples from the respective category out of the total number of samples from non-archival studies published within the period. Findings from time-based regression analyses: ***—significant decrease (−) ($p < 0.01$); **—significant decrease (−) ($p < 0.05$). *Panel B:* [c] Absolute number of samples involving professionals from the respective category per research method. [d] Percentage of samples involving professionals from the respective category out of the total number of samples from the respective research method. Chi-square test for cross-tabulation of research methods and primary occupation: $\chi^2 = 253.889$; $p < 0.001$; n = 318.

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

based regressions reveal significant decreases in studies involving managers below the top-management level and studies that do not disclose the occupation of their participants ($p < 0.01$ and $p < 0.05$ respectively).

Table 8 (Panel B) reports the cross-tabulation of research methods and the primary occupations of participants. Most notably, 50.0% (45.5%) of the case studies (field studies) involve multiple professional groups. Untabulated findings suggest that the share of studies involving multiple professional groups increases for case and field studies over time. Given that the consideration of different perspectives on an issue may facilitate the generalization of findings among individuals, we consider this tendency an improvement of external validity in the context of case and field studies. We also detect an increasing involvement of multiple groups for survey research. This finding appears notable because survey studies frequently imply a key information bias. This bias emerges when surveys rely on responses by one member who reports on organizational issues and may generalize his or her individual perceptions to the entire organization (Kumar et al., 1993). This issue may be addressed by querying multiple informants (Homburg et al., 2012). However, most of the survey studies that involve multiple professional groups still refer to one respondent per organization. We assigned them to the "Multiple" category because the professional backgrounds of the key informants differ among organizations. This finding not only implies that key informant bias continues to be an issue for most survey research. In addition, occupational differences that are not controlled for may increase error variance and decrease the power of statistical tests (Shadish et al., 2002, p. 51). Against this background, the decreasing share of experiments that involve multiple professional groups constitutes an improvement. In addition, we find that graduate students and graduates increasingly participate in experiments.

Overall, the broader range of settings considered and the increasing involvement of different professional groups in case and field studies synoptically provide some support for the improvement of external validity, as suggested by Proposition 5. However, considerable threats to the validity of survey studies persist.

4.2.3. Construct validity (Proposition 6)

Proposition 6 suggests that construct validity increases over time. To illuminate this issue, we rely on the numbers of measures for construct validation (Code 8) and reliability (Code 9). In addition, we refer to the type of dependent variables (Code 10) and the numbers of data sources (Code 11) on which the studies draw (Scandura and Williams, 2000). By referring to the numbers of actions taken to address construct validity and reliability reported in the papers, we implicitly assume that these measures contribute to improved construct validity and reliability. Table 9 (Panel A) provides descriptive statistics on the number of measures for construct validation. We find that the mean values of the reported measures increase over time. Multi-sample median and Kruskal-Wallis tests indicate significant differences regarding the median and distribution of the measures for construct validation ($p < 0.05$) and thus corroborate Proposition 6. In particular, we detect an increase in the employment of factor analysis and tests on discriminant, convergent or predictive validity among quantitative studies. Case and field studies increasingly apply data triangulation and chains of evidence.

Table 9 (Panel B) suggests a similar tendency for the number of reliability measures (Code 9). However, in contrast to Code 8, we do not find a significant increase in the number of measures taken. Nonetheless, a Kruskal-Wallis test indicates a significant difference for the distribution of the numbers of measures ($p < 0.05$). For instance, computation of Cronbach's alpha and the compilation of case study protocols are frequently applied to assess reliability. Based on the assumptions that more extensive reports on the measures taken imply improvements, our analyses largely corroborate Proposition 6.

Code 10 refers to the types of dependent variables employed in PMAR. In particular, unverified self-reports may compromise the validity of the findings due to subjective biases (Van der Stede et al., 2005). Table 10 reports the findings derived from the category scheme by Scandura and Williams (2000). Subjective performance evaluations, attitudinal outcomes and perceptual outcomes mainly consist of self-reports that may introduce bias (Birnberg et al., 1990). Our findings indicate that such data are decreasingly collected over time. Instead, PMAR relies increasingly on tangible outcomes in terms of objective measures. Time-based regressions reveal that the increase (decrease) in the reliance on tangible outcomes (attitudinal outcomes) is significant at the 1% level. Given that objective measures are not subject to reactivity effects and subjective biases (Birnberg et al., 1990), we consider these developments an increase in construct validity.[14]

Because the employment of multiple data sources may improve construct validity (Modell, 2005), we eventually illuminate whether the studies collect data from a single or multiple sources (Code 11). The findings reported in Table 11 (Panel A) show that the share of studies involving multiple sources increases over time. However, even in the 2010s, 55.0% of the studies still rely on a single source of data. Time-based regressions suggest continuity in the development as they reveal a significant increase at the 1% level.

Table 11 (Panel B) presents the findings on the association between the sources of data and research methods. This cross-tabulation seems important because an increasing reliance on multiple sources may be attributable to particular research methods. For instance, case and field studies frequently involve different sources of data, such as interviews, observations or the analysis of internal documents (Ahrens and Chapman, 2006; Merchant and Van der Stede, 2006). In contrast, survey research is often subject to a common method bias. This bias arises when independent and dependent variables are derived from survey responses by the same person (Van der Stede, 2014). Table 11 (Panel B) suggests that the increasing share of studies that involve multiple sources of data may be primarily attributed to archival (particularly in the North American journals) as well as case and field studies (particularly in the European journals). Survey research mostly draws on one source of data (80.9%). In fact, the share of survey studies employing multiple sources of data even declines over time. Against this background, we conclude that an increasing reliance on different sources of data – which we consider an improvement in construct validity – does not pertain to all research methods. In summary, we find considerable support for Proposition 6; however, not all coding dimensions indicate virtual improvements.

4.2.4. Statistical conclusion validity (Proposition 7)

According to Proposition 7, statistical conclusion validity increases over time. For this reason, we analyse whether authors increasingly test for specific threats to statistical conclusion validity (Code 12). In this context, we place particular emphasis on endogeneity issues that bias statistical associations between variables and may lead to incorrect conclusions about their factual relationship (Evans et al., 2015). Endogeneity may be the consequence of omitted correlated variables, simultaneity or self-selection

[14] A comparison of North American and European journals indicates that this tendency is primarily driven by papers published in North American journals, which increasingly employ tangible outcomes (34.4% on overage among the periods). In addition, these papers strongly rely on behavioural outcomes (27.6%). By contrast, the types of dependent variables most frequently employed in papers in European journals are subjective performance evaluations (21.7%) and perceptual outcomes (26.1%).

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

Table 9
Descriptive statistics on reported measures for construct validation and reliability.

	Panel A: Number of measures for construct validation (Code 8)			Panel B: Number of reliability measures (Code 9)		
Period	n	Mean	Std. dev.	n	Mean	Std. dev.
1980–1982	25	0.72	0.843	25	0.48	0.714
1990–1992	68	0.74	0.987	74	0.41	0.571
2000–2002	103	1.16	1.109	108	0.60	0.610
2010–2012	142	1.58	1.460	166	0.64	0.613
Total period	338	1.22	1.277	373	0.59	0.627

Refers to studies for which procedures that ensure construct validity and reliability respectively are de facto applicable. For instance, archival studies that do not include complex measurement constructs but mainly refer to data available in databases e.g., compensation data, were assigned to the category "not applicable", which is not part of the analyses presented in this table. Multi-sample median and Kruskal-Wallis tests indicate significant differences regarding median and distribution of the number of measures for construct validation at the 5% level. The same procedures indicate a significant difference with regard to the distribution of the numbers of the reliability measures at the 5% level.

Table 10
Types of dependent variables.

Type of dependent variables (Code 10) per period						
	1980–1982 # (%)	1990–1992 # (%)	2000–2002 # (%)	2010–2012 # (%)	Total	Over-all trend
Subjective performance evaluations	4[a] (18.2%)[b]	12 (21.4%)	22 (26.5%)	20 (15.2%)	58	
Tangible outcomes	2 (9.1%)	16 (28.6%)	28 (33.7%)	50 (37.9%)	96	***(+)
Behavioural outcomes	11 (50.0%)	19 (33.9%)	21 (25.3%)	37 (28.0%)	88	
Attitudinal outcomes	9 (40.9%)	9 (16.1%)	10 (12.0%)	5 (3.8%)	33	***(−)
Perceptual outcomes	5 (22.7%)	9 (16.1%)	22 (26.5%)	32 (24.2%)	68	
Other	1 (4.5%)	2 (3.6%)	8 (9.6%)	10 (7.6%)	21	

Refers to the number of samples that comprise analyses of relationships between dependent and independent variables. Qualitative studies as well as quantitative descriptive studies are excluded (n = 293). [a] Absolute number of samples employing the respective type of variable as dependent variable. [b] Percentage of samples employing the respective type of dependent variable out of the total number of samples comprising analyses of relationships between dependent and independent variables (qualitative and quantitative descriptive studies excluded) published within the period. Note that one study may employ different types of dependent variables and may be assigned to multiple categories. As a consequence, the percentages do not total to 1. Findings from time-based regression analyses: ***—significant increase (+) or decrease (−) ($p < 0.01$).

Table 11
Numbers of data sources.

Panel A: Number of data sources (Code 11) per period						
	1980–1982 # (%)	1990–1992 # (%)	2000–2002 # (%)	2010–2012 # (%)	Total	Over-all trend
Single	20[a] (80.0%)[b]	56 (68.3%)	75 (61.0%)	94 (55.0%)	245	***(−)
Multiple	4 (16.0%)	22 (26.8%)	46 (37.4%)	77 (45.0%)	149	***(+)
No report	1 (4.0%)	4 (4.9%)	2 (1.6%)	0 (0.0%)	7	***(−)

Panel B: Number of data sources (Code 11) per research method						
	Archival	Survey	Experimental	Case study	Field Study	Total
Single	21[c] (25.3%)[d]	89 (80.9%)	105 (91.3%)	6 (15.8%)	24 (43.6%)	245
Multiple	62 (74.7%)	21 (19.1%)	10 (8.7%)	29 (76.3%)	27 (49.1%)	149
No report	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (7.9%)	4 (7.3%)	7

Refers to the number of samples (n = 401). *Panel A:* [a] Absolute number of samples. [b] Percentage of samples from the respective category out of the total number of samples published within the period. Findings from time-based regression analyses: ***—significant increase (+) or decrease (−) ($p < 0.01$). *Panel B:* [c] Absolute number of samples relying on the respective sources of data per research method. [d] Percentage of samples relying on the respective sources of data per research method out of the total number of samples from the respective research method. Chi-square test for cross-tabulation of research methods and sources of data: $\chi^2 = 164.399$; $p < 0.001$; n = 401.

(Chenhall and Moers, 2007; Nikolaev and van Lent, 2005). For this reason, we investigate whether statistical procedures, such as two stage least square estimations or propensity score analyses, are performed to control for endogeneity concerns. We also consider whether further issues that may bias statistical associations, such as multicollinearity, heteroscedasticity or outliers, are considered.

Table 12 reports the findings on this coding dimension. Because the aforementioned threats primarily apply to quantitative non-experimental research designs (Chenhall and Moers, 2007; Ittner, 2014), this analysis refers to archival and survey research only. The findings reveal two noteworthy patterns. On the one hand, we find that even in the 2010s only a minor proportion of archival and survey papers address endogeneity issues. For instance, the omission of correlated variables is only considered in 21.4% of the studies, whereas only 15.5% of the studies control for reverse causality (simultaneity). On the other hand, we find a strong increase in testing for all of these threats. Time-based linear regressions emphasize this development as we detect significant increases over time ($p < 0.01$). Therefore, our findings are in line with Proposition 7, but still indicate considerable potential for improvement.

5. Discussion

This paper assesses diversity and validity in PMAR based on a content analysis of 375 papers published over four decades. We argue that a diverse set of research methods and theories, along with the consideration of validity, constitute important prerequisites for the accumulation of knowledge on MA practice (e.g., Bonner et al., 2012; Edwards et al., 2013; Hoque et al., 2013; Luft and Shields, 2002; Merchant et al., 2003). Our analysis relies on the propositions that the diversity of research approaches and the validity of inferences increase as PMAR moves forward from the description of MA practice towards its explanation. However, in light of recent debates on the state of MA research, we consider

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

Table 12
Consideration of threats to statistical conclusion validity.

Consideration of threats to statistical conclusion validity (Code 12) per period						
	1980–1982 # (%)	1990–1992 # (%)	2000–2002 # (%)	2010–2012 # (%)	Total	Over-all trend
Test for multicollinearity	0[a] (0.0%)[b]	7 (17.5%)	14 (22.2%)	42 (50.0%)	63	***(+)
Test for omitted variable bias	0 (0.0%)	1 (2.5%)	11 (17.5%)	18 (21.4%)	30	***(+)
Test for simultaneity bias	0 (0.0%)	0 (0.0%)	3 (4.8%)	13 (15.5%)	16	***(+)
Test for self-selection bias	2 (33.3%)	4 (10.0%)	16 (25.4%)	29 (34.5%)	51	***(+)
Test for heteroscedasticity	0 (0.0%)	3 (7.5%)	10 (15.9%)	43 (51.2%)	56	***(+)
Test for outliers	0 (0.0%)	4 (10.0%)	13 (20.6%)	25 (29.8%)	42	***(+)
Assertation of model identification	0 (0.0%)	0 (0.0%)	2 (3.2%)	3 (3.6%)	5	

Refers to the number of samples that rely on quantitative non-experimental research designs (archival and survey studies; n = 193). [a] Absolute number of samples employing a procedure to test for the respective threat. [b] Percentage of samples from the respective category out of the total number of samples employing quantitative non-experimental research design published within the period. Findings from time-based regression analyses: ***—significant increase (+) ($p < 0.01$).

that institutional pressures may put such progress into perspective (e.g., Birnberg, 2009; Merchant, 2010).

5.1. Conclusions on the development of diversity in PMAR

Our findings on *diversity* provide limited evidence of an increasing narrowness of PMAR over time. More precisely, our diversity-related analyses allow for two major conclusions. On the one hand, PMAR has become narrower in terms of topics as it increasingly addresses control issues. On the other hand, we find that PMAR – unlike financial accounting research (e.g., Demski, 2007; Lukka, 2010; Moser, 2012) – represents a highly diversified discipline with respect to research methods and theoretical perspectives. Taken together, the decreasing content diversity and the high degrees of method and theory diversity may imply that PMAR is losing the "faddish nature" and descriptive character for which it has been criticized in the past (Ittner and Larcker, 2001; Zimmerman, 2001). Given that Lukka and Mouritsen (2002, p. 806) consider a declining openness to different research approaches "regression, not progress, within the management accounting academia", the sustained diversity in methods and theories seems particularly notable. It constitutes an important premise for the protection of a "hallmark" (Krishnan, 2015, p. 182) and the further progress of the discipline (Luft and Shields, 2002; Merchant et al., 2003).

Therefore, we conclude that the current state of PMAR does not appear as troubling as indicated by previous commentaries (e.g., Birnberg, 2009; Lukka, 2010; Merchant, 2013). However, our findings raise a number of questions. First, assuming that PMAR should be relevant to practice (e.g., Baldvinsdottir et al., 2010; Krishnan, 2015; Malmi and Granlund, 2009; Mitchell, 2002; Modell, 2014), the question arises whether the increasing focus on control issues implies practical relevance. Future research might illuminate whether MA practice is in need of improvement of control systems or whether decreasing content diversity is a consequence of an increasing self-referentiality of MA research (Birnberg, 2009; Shields, 1997; Van der Stede, 2015). The latter might indicate a growing research-practice gap (Hopwood, 2008b; Kaplan, 2011; Tucker and Parker, 2014). In this context, the recent decrease in papers on other topics appears particularly noteworthy. For instance, digitization and big data, risk management and sustainability accounting constitute challenges for MA practice (Bhimani and Willcocks, 2014; Unerman and Chapman, 2014; Van der Stede, 2011) that are potentially under-researched by PMAR. If such topics are addressed by researchers from other disciplines instead (Kaplan, 2011; Van der Stede, 2015), the relevance of PMAR might decline. Practitioner surveys (e.g., Foster and Young, 1997) thus might help to explore further avenues of research that imply potential for "research-for-practice-sake" (Birnberg, 2009; p. 6).

Second, a further narrowing of research topics might be reinforced when the significant increase in archival studies continues. Anecdotal evidence suggests that the availability of archival data has narrowed the focus of financial accounting research (Hopwood, 2007; Kaplan, 2011). Employing archival data implies a series of advantages. In particular, such data avoids the issues related to self-reports that confine construct validity (see Section 4.2.3 and Birnberg et al., 1990; Van der Stede et al., 2005). Moreover, the growing share of archival studies may reflect new research opportunities because regulatory changes require disclosure of data relevant to MA researchers (Van der Stede, 2011; Wagenhofer, 2016). However, increasing reliance on publicly available data may imply a self-selection of research topics. In other words, issues with available data might be preferred over issues that are most relevant to practice. Still, this threat is put into perspective as long as the leading MA journals follow their established path of openness regarding method and theory diversity and encourage submissions that rely on different research approaches. This reasoning applies particularly to qualitative approaches, for which our findings suggest a pivotal role in the exploration of emerging topics.

Third, our supplementary analyses on European and North American journals reveal two groups of journals with distinctive patterns concerning research methods and theories. Although favourable in terms of method and theory diversity, the different emphases may indicate a further segregation in the MA research community identified by previous studies (e.g., Hesford et al., 2007; Lukka and Granlund, 2002). Based on citations in AOS, CAR, JAE, JAR and TAR, Bonner et al. (2012) conclude that MA researchers cite one another irrespective of the methods or theories applied. However, it remains an open question whether such communication occurs also between North American and European journals beyond AOS. If there were segregation, PMAR would not exploit the full potential of method and theory diversity because findings from one perspective might not be considered in research that provides a complementary perspective. We argue that these questions provide avenues for research to deepen our understanding of MA knowledge production.

5.2. Conclusions on the development of validity in PMAR

Our findings on *validity* partially correspond with our propositions that all four types of validity increase over time. While some developments do not appear continuous, others still indicate considerable potential for improvement. However, we conclude that most criteria are increasingly addressed over time. For instance, the increasing number of measures for construct validation suggests that researchers pay greater attention to the operationalization of theoretical constructs. Moreover, we consider the stronger reliance on objective measures as dependent variables an improvement compared to findings from previous analyses (Van der Stede et al., 2005). We also detect a recent increase in internal validity as the share of longitudinal studies has grown in the 2010s. Moreover, Van der Stede (2014) states that "there is little that can be done

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

to substantially enhance the confidence in causal statements [. . .] beyond [. . .] having good, compelling theory" (p. 570). Against this background, the declining share of studies that do not rely on an explicit theoretical perspective suggests a further improvement with regard to internal validity. Ultimately, our findings suggest an increasing consideration of threats to statistical conclusion validity, indicating that the issue of biased associations and corresponding incorrect statistical conclusions declines. Such improvements may appear somewhat self-evident if we expect that science not only accumulates new knowledge but also advances the techniques on which it relies. However, Scandura and Williams (2000), for instance, find that the validity of organizational studies deteriorates over time. Thus, the aforementioned improvements are obviously less a matter of course.

An additional reflection on our findings indicates that individual research methods contribute to the increase in validity to varying degrees. In particular, both case and field studies imply considerable improvements to internal validity due to an increasing reliance on longitudinal research designs. In addition, the escalating involvement of multiple professional groups in case and field studies implies an increase in external validity. The illumination of MA issues from different perspectives suggests that findings may be generalized across different occupational groups. These substantial contributions to validity increases appear remarkable in light of past criticism that qualitative research approaches are too descriptive and contribute little to theory building (Zimmerman, 2001).

Regarding experimental research, we find that graduate students and graduates increasingly participate in these studies. This tendency might improve the external validity of findings on more complex tasks, provided that the decisions and perceptions of participants with greater work experience and MA knowledge are better proxies for decision-making and perceptions in corporate practice. Moreover, we find that experimental research increasingly relies on homogeneous samples, which are likely to increase the power of statistical tests. Archival studies contribute to internal validity as they increasingly employ longitudinal data and to construct validity as they typically use objective measures as dependent variables. However, endogeneity issues are increasingly, yet still rarely, addressed.

In contrast to these improvements, survey studies have progressed only with regard to the number of measures for construct validation and with respect to threats to statistical conclusion validity. Data are still almost exclusively collected at a single point in time. Given that cross-sectional surveys do not provide evidence of causality (Van der Stede, 2014), potential for improvement persists. Statistical conclusions are still frequently threatened due to rare testing for the omission of correlated variables, simultaneity and self-selection. Moreover, most survey research still involves one key informant per organization. Van der Stede (2014) suggests that surveys may use multiple respondents or multiple sources of data to improve the validity of findings. For this reason, survey research that employs a dyadic research design – such as a management accountant that provides information on particular MA approaches employed and a manager that reports on his or her perceptions of MA tools – may represent an avenue for the further methodological advancement of PMAR. Following this path appears particularly important because PMAR focuses heavily on the organizational level. Although PMAR obviously first intended to understand how MA affects individuals, it increasingly refers to the implications of MA for organizational behaviour. Given the increase in tangible outcomes as dependent variables, this finding may imply that PMAR pays increasing attention to the impact of MA on organizational performance. The literature stresses that illuminating the link between MA choices and performance is particularly valuable from a practitioners' perspective (e.g., Ittner and Larcker, 2002; Malmi and Granlund, 2009). However, the validity of such findings is threatened if researchers make inferences from one organizational member to the entire organization. Although Van der Stede et al. (2005) addressed this issue a decade ago, it still seems to weigh heavily because of the strong focus of PMAR on the organizational level.

Our analyses on validity reveal a series of additional insights into the scope of PMAR. Our investigation of the type of samples indicates that researchers are devoting more attention to the public sector. We argue that this tendency may be a response to recent new public management initiatives (e.g., Jacobs et al., 2004; Llewellyn and Northcott, 2005). Similarly, the finding that top-level managers increasingly participate in MA studies may mirror failures in recent corporate governance practices (e.g., Drymiotes, 2011). Despite the aforementioned uncertainty whether PMAR addresses current issues in MA practice, such findings provide a first indication that PMAR reflects recent developments in practice after all.

Eventually, the limited involvement of management accountants highlighted in our analysis of primary occupations appears puzzling. These accountants are typically responsible for the design of MA tools and procedures (Kaplan, 2006). Therefore, obtaining information from management accountants may lead to insights into the technical specifics and determinants of MA systems. Instead, PMAR has clearly moved from focusing on technical details to analysing its social implications (Baldvinsdottir et al., 2010). However, Baldvinsdottir et al. (2010) argue that an adequate exploration of MA implications requires a thorough understanding of the techniques employed. Therefore, greater emphasis on MA techniques may deepen our understanding and contribute to the further advancement of MA practice. Likewise, MA practice may benefit from research that interlinks the effects of MA techniques at the individual as well as organizational level. Our investigation of the levels of analysis suggests that previous research has focused on either. However, the simultaneous investigation may lead to more comprehensive insights and thus may prove beneficial in guiding MA choices.

5.3. Limitations and outlook

Our study is subject to some limitations. We acknowledge that our non-random sampling impedes inferences regarding other years. Therefore, our study provides a comparison of four specific periods. We further acknowledge that our study focuses on major outlets for the publication of PMAR. The journal selection is not exhaustive and may be complemented in several ways. Although we consider the simultaneous investigation of different research methods helpful for a comprehensive overview of the state and path of PMAR, this approach prevents a more detailed investigation of the peculiarities of individual research methods. Our analysis – derived from the framework by Scandura and Williams (2000) – may be complemented by more specific methodological reviews that provide an "ex post" perspective on individual research methods and employ a larger range of indicators (similar to Van der Stede et al. (2005) on survey research). Regardless of these limitations, we expect our study to be useful to MA researchers as an overview and point of departure for debates on the future path of PMAR. In particular, our study provides empirical evidence on issues that have been heavily discussed in recent years but have mostly relied on anecdotal evidence. Most importantly, we consider our findings an encouragement to MA researchers to follow the path of diversity to distance PMAR from related accounting disciplines that are claimed to be narrower.

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

Appendix A

Coding dimensions concerning diversity (Codes 1–3)		
Code 1	Content	
Description:	Refers to the subject areas studied and reflects content diversity. Relies on a slightly modified version of the categorization scheme by Hesford et al. (2007), but excludes the sub-category "Accounting Information Systems" as it may be considered a separate accounting sub-discipline (e.g., DeFond, 2015).	
Subject of coding:	Each paper (n = 375)	
Category 1	Cost and planning	Studies on cost variance analysis or the allocation of overhead costs as well as planning issues. Sub-categories: (1) Cost allocation; (2) Other cost accounting topics; (3) Cost practices; (4) Budgeting; (5) Capital budgeting; (6) Multiple.
Category 2	Control	Studies on performance measurement and control. Sub-categories: (7) Performance measurement and evaluation; (8) Organizational control; (9) International control; (10) Multiple.
Category 3	Other	Applies to papers that cannot be subsumed under the first two categories. Sub-categories: (11) Benchmarking; (12) Quality (TQM); (13) Just-in-time; (14) Research methods; (15) Strategic management; (16) Transfer pricing; (17) Multiple.
Code 2	Research method	
Description:	Refers to the methods employed and relies on a slightly modified version of the categorization scheme developed by Hesford et al. (2007) by excluding non-empirical categories, reflecting method diversity.	
Subject of coding:	Each sample (n = 401); for papers that comprise more than one sample, we considered each sample a separate study.	
Category 1	Archival	Analysis of secondary archival data (e.g., data on executive compensation) with quantitative methods (Moers, 2007).
Category 2	Survey	Data collection based on standardized questionnaires that are disseminated by mail, e-mail or online (Birnberg et al., 1990; Van der Stede et al., 2005).
Category 3	Experimental	Investigations involving human subjects completing a task in an artificial setting, along with the manipulation and corresponding measurement of variables (Lindquist and Smith, 2009). Includes also field experiments that take place in the natural setting of the participating subjects (e.g., an organizational context) (Birnberg et al., 1990).
Category 4	Case study	Investigations within a single organization, employing data sources such as interviews, internal documents and observations (Birnberg et al., 1990; Hesford et al., 2007; Lindquist and Smith, 2009).
Category 5	Field study	Investigations of more than one site employing data sources such as interviews, internal documents and observations (Birnberg et al., 1990; Hesford et al., 2007; Lindquist and Smith, 2009).
Code 3	Theoretical perspective	
Description:	Refers to the theoretical perspectives adopted and reflects theory diversity. Relies on the categorization scheme by Shields (1997), but excludes the category "Strategic management" as such studies usually rely on theories from the other perspectives (e.g., organizational behaviour or economics).	
Subject of coding:	Each sample (n = 401); for papers that comprise more than one sample, we considered each sample a separate study.	
Category 1	Economics	e.g., studies informed by agency theory, transaction cost theory.
Category 2	Psychology	e.g., studies informed by motivation, cognitive or social psychology theories.
Category 3	Sociology	e.g., studies informed by institutional theory, contingency theory, structuration theory.
Category 4	Organizational behaviour	e.g., studies informed by role theory, network theory, organizational change theories.
Category 5	Production and operations management	Studies informed by theories related to technology, quality or production layout.
Category 6	Multiple	Studies informed by multiple theoretical perspectives.
Category 7	No theory	Studies without a recognizable explicit or implicit reference to a theoretical perspective.

Supplementary dimension of analysis (Code 4)		
Code 4	Level of analysis	
Description:	Refers to the level of analysis primarily focused by each study and relies on the levels distinguished by Luft and Shields (2003) and Van der Stede et al. (2005). The level of analysis refers to the dependent and independent variables employed in a study and is in accordance with Luft and Shields (2003, p. 175) defined as "the level at which the variation of interest occurs". Studies referring to more than one level of analysis are assigned to category 5 ("cross-level"). If no variables are employed, the level of analysis is based on the level to which the research questions refer to.	
Subject of coding:	Each sample (n = 401); for papers that comprise more than one sample, we considered each sample a separate study.	
Category 1	Individual	e.g., studies on the impact of performance measures on managerial performance.
Category 2	Subunit	e.g., studies on the impact of performance measures on business unit performance.
Category 3	Organization	e.g., studies on the impact of performance measures on firm performance.
Category 4	Beyond organization	e.g., studies on the use of cost accounting information in supply chains.
Category 5	Cross-level	Studies referring to different levels of analysis (e.g., managerial motivation as well as organizational performance).

Coding dimension concerning internal validity (Code 5)		
Code 5	Time frame	
Description:	Affects internal validity as longitudinal studies allow the researcher to directly identify causal relationships, in which causes precede the effect (Scandura and Williams, 2000; Van der Stede, 2014).	
Subject of coding:	Each sample (n = 401); for papers that comprise more than one sample, we considered each sample a separate study.	
Category 1	Cross-sectional	Data-gathering at one point in time.
Category 2	Longitudinal	Data-gathering over a time period; i.e., data is collected at least at two different points in time. Given that the manipulation of variables in experimental studies precedes the observation of the dependent variables, we considered experimental research quasi-longitudinal and assigned them to this category.

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

Coding dimensions concerning external validity (Codes 6 and 7)		
Code 6	Type of sample	
Description:	Affects external validity as investigations of different settings allow generalizing findings among different contexts. Refers to the scheme by Scapens and Bromwich (2010b) with the addition of "Not-for-profits" as another category (Scandura and Williams, 2000).	
Subject of coding:	Each sample (n = 401); for papers that comprise more than one sample, we considered each sample a separate study.	
Category 1	Generic	Studies in a laboratory setting.
Category 2	Manufacturing	Studies on companies operating in different manufacturing industries.
Category 3	Specific industries	Studies on companies operating in particular manufacturing industries.
Category 4	Services	Studies on companies operating in the service sector.
Category 5	Public sector	Studies on governmental entities.
Category 6	Not-for-profit	Studies on nongovernmental, non-profit-making entities.
Category 7	Specific countries	Studies on MA in a particular country or region.
Category 8	Multinationals	Studies on MA specifics in multinational companies.
Category 9	Other	Primarily comprises studies on companies operating in manufacturing or service industries.
Code 7	Primary occupation	
Description:	Affects external validity as the involvement of participants with different backgrounds may enable generalizing findings among different individuals, groups and occupations. Refers to the modified scheme by Scandura and Williams (2000). If laboratory experiments involve MBA students and their occupations were mentioned, they were assigned to one of the first four categories. When graduates and undergraduates were involved, we coded the study as "Multiple".	
Subject of coding:	Each sample (n = 401); for papers that comprise more than one sample, we considered each sample a separate study. If studies exclusively employ an archival research design and do not involve individuals, they were assigned to category 9.	
Category 1	Management accountants and accountants	Studies involving employees devising and operating management accounting systems.
Category 2	Top-level management	Studies involving managers at the top level of a firm's hierarchy (e.g., CEO, CFO).
Category 3	Other management	Studies involving managers below the top level (e.g., profit center managers).
Category 4	Others	Studies involving, e.g., auditors, health care professionals, engineers, deans of academic departments.
Category 5	Multiple	Studies involving participants with different occupational backgrounds, e.g., management accountants and managers.
Category 6	Undergraduate students	Laboratory experiments involving entry level students, usually obtaining a bachelor's degree.
Category 7	Graduate students/graduates	Laboratory experiments involving graduate students (e.g., MBA students) and graduates without further specification of their profession.
Category 8	No report	Applies to studies that do not provide information on the subjects involved.
Category 9	Not applicable	Applies to studies that rely exclusively on archival data.

Coding dimensions concerning construct validity (Codes 8–11)		
Code 8	Number of measures for construct validation	
Description:	Reflects which measures (i.e., courses of action) have been taken to address construct validity. As we refer to the corresponding information provided in the papers, we analyse "disclosed validity". Relies on the category scheme by Scandura and Williams (2000) for quantitative research designs and for qualitative research on suggestions by Lillis (1999), Modell (2005) and Yin (2014). As we refer to the number of measures taken, we implicitly assume that construct validity increases with the number of measures taken.	
Subject of coding:	Each sample (n = 401); for papers that comprise more than one sample, we considered each sample a separate study.	
Category 1	Quantitative research designs	Procedures applicable to quantitative research designs (Modell, 2005; Scandura and Williams, 2000). Sub-categories: (1) Confirmatory factor analysis; (2) Exploratory factor analysis; (3) Reports on discriminant validity, convergent validity and predictive validity; (4) Method triangulation.
Category 2	Qualitative research designs	Procedures applicable to qualitative research designs (Lillis, 1999; Yin, 2014). Sub-categories: (5) Multiple sources of evidence/data triangulation; (6) Method triangulation; (7) Chain of evidence; (8) Case study report/transcripts/field notes/draft reviewed by key informants; (9) Analytical protocol/content analysis; (10) Coding by multiple researchers/intercoder reliability.
Category 3	No report	Applies to studies that do not provide information on the measures taken to ensure construct validation.
Category 4	Not applicable	Applies to studies for which procedures that ensure construct validity are not applicable (e.g., archival studies), which do not include complex measurement constructs and primarily refer to data available in databases.
Code 9	Number of reliability measures	
Description:	Refers to measures (i.e., courses of action) that have been taken to ensure the reliability of the research designs employed. Similarly to Code 8, we implicitly assume that reliability increases with the number of measures taken. Relies on approaches suggested by Drost (2011) for quantitative research designs and for qualitative research on suggestions by Yin (2014).	
Subject of coding:	Each sample (n = 401); for papers that comprise more than one sample, we considered each sample a separate study.	
Category 1	Quantitative research designs	Procedures that assess the reliability in the context of quantitative research designs (Drost, 2011). Sub-categories: (1) Internal consistency (Cronbach's Alpha); (2) Alternative forms of measures; (3) Test-retest reliability; (4) Split-half approach; (5) Interrater reliability.
Category 2	Qualitative research designs	Procedures that contribute to the reliability in the context of qualitative research designs (Yin, 2014). Sub-categories: (6) Case study protocol; (7) Case study database (consisting of material collected during field visits, e.g., field notes, internal documents, narratives).

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

Category 3	No report	Applies to studies that do not provide information on the measures taken to ensure reliability.
Category 4	Not applicable	Applies to studies for which procedures that ensure reliability are not applicable.
Code 10	Type of dependent variables	
Description:	Reflects construct validity as particular types of dependent variable may induce measurement errors or may imply noise (e.g., self-reports that may be subject to subjective biases) (Birnberg et al., 1990; Van der Stede et al., 2005). Relies on a refined category scheme by Scandura and Williams (2000).	
Subject of coding:	Each sample (n = 401); for papers that comprise more than one sample, we considered each sample a separate study. Case and field studies and quantitative descriptive studies were assigned to Category 7. Multiple coding applies when studies involve several dependent variables.	
Category 1	Subjective performance evaluations	Subjective assessments of a company's or a sub-unit's performance (mostly in comparison to its competitors), subjective assessment of one's own managerial performance, task performance.
Category 2	Tangible outcomes	e.g., turnover or Return on Investment.
Category 3	Behavioral outcomes	e.g., decisions made.
Category 4	Attitudinal outcomes	e.g., pay satisfaction.
Category 5	Perceptual outcomes	e.g., perceived fairness.
Category 6	Other	e.g., propensity that a particular MA approach gets implemented.
Category 7	Not applicable	Applies to studies that do not investigate associations and do not explicate dependent variables.
Code 11	Number of data sources	
Description:	The employment of multiple data sources may allow validation of findings by overcoming limitations of individual data sources. For this reason, we implicitly assume that employing more than one data source implies higher construct validity.	
Subject of coding:	Each sample (n = 401); for papers that comprise more than one sample, we considered each sample a separate study.	
Category 1	Single	Studies that rely on one data source, e.g., a survey employing a written questionnaire that is completed by one respondent per organization.
Category 2	Multiple	Studies that rely on at least two data sources, e.g., interviews and internal documents in a field study.
Category 3	No report	Applies to studies that do not provide information on the number of data sources.

Coding dimension concerning statistical conclusion validity (Code 12)

Code 12	Consideration of threats to statistical conclusion validity	
Description:	Refers to measures (i.e., particular tests) that control for threats that may lead to incorrect conclusions based on statistical analysis. Relies on issues discussed by Chenhall and Moers (2007), Evans et al. (2015), Nikolaev and van Lent (2005) and Wooldridge (2013).	
Subject of coding:	Each sample (n = 401); for papers that comprise more than one sample, we considered each sample a separate study.	
Category 1	Test for multicollinearity	e.g., computation of variance inflation factors.
Category 2	Test for omitted variable bias	e.g., instrumental variable estimation.
Category 3	Test for simultaneity bias	e.g., two-stage least squares estimation.
Category 4	Test for self-selection bias	e.g., Heckman correction.
Category 5	Test for heteroscedasticity	e.g., estimation of heteroscedasticity-robust standard errors.
Category 6	Test for outliers	e.g., winsorizing of data.
Category 7	Assertion of model identification	e.g., statement on how a parameter identification problem has been addressed.
Category 8	Not applicable	Applies to qualitative studies, which do not use statistics for inferences about relationships under study, and experimental studies, which establish causal linkages between variables due to experimental control (i.e., randomization of treatments, manipulation of variables).

References

Abernethy, M.A., Chua, W.F., Luckert, P.F., Selto, F.H., 1999. Research in managerial accounting: learning from others' experiences. Account. Finance 39, 1–27.
Ahrens, T., Chapman, C.S., 2006. Doing qualitative field research in management accounting: positioning data to contribute to theory. Organ. Soc. 31, 819–841.
Ahrens, T., Dent, J.F., 1998. Accounting and organizations: realizing the richness of field research. J. Manage. Account. Res. 10, 1–39.
Atkinson, A.A., Shaffir, W., 1998. Standards for field research in management accounting. J. Manage. Account. Res. 10, 41–68.
Baldvinsdottir, G., Mitchell, F., Nørreklit, H., 2010. Issues in the relationship between theory and practice in management accounting. Manage. Account. Res. 21, 79–82.
Ballas, A., Theoharakis, V., 2003. Exploring diversity in accounting through faculty journal perceptions. Contemp. Account. Res. 20, 619–644.
Bhimani, A., Willcocks, L., 2014. Digitisation, 'Big Data' and the transformation of accounting information. Account. Bus. Res. 44, 469–490.
Bhimani, A., 2002. European management accounting research: traditions in the making. Eur. Account. Rev. 11, 99–117.
Birnberg, J.G., Shields, M.D., Young, S.M., 1990. The case for multiple methods in empirical management accounting research (with an illustration from budget setting). J. Manage. Account. Res. 2, 33–66.
Birnberg, J.G., 2009. The case for post-modern management accounting: thinking outside the box. J. Manage. Account. Res. 21, 3–18.
Bisbe, J., Batista-Foguet, J.-M., Chenhall, R., 2007. Defining management accounting constructs: a methodological note on the risks of conceptual misspecification. Account. Organ. Soc. 32, 789–820.
Blau, P.M., 1977. Inequality and Heterogeneity: A Primitive Theory of Social Structure. Free Press, New York and London.
Bonner, S.E., Hesford, J.W., Van der Stede, W.A., Young, S.M., 2006. The most influential journals in academic accounting. Accounting. Organ. Soc. 31, 663–685.
Bonner, S.E., Hesford, J.W., Van der Stede, W.A., Young, S.M., 2012. The social structure of communication in major accounting research journals. Contemp. Account. Res. 29, 869–909.
Brutus, S., Aguinis, H., Wassmer, U., 2013. Self-reported limitations and future directions in scholarly reports: analysis and recommendations. J. Manage. 39, 48–75.
Callen, J.L., 2015. A selective critical review of financial accounting research. Crit. Perspect. Account. 26, 157–167.
Chapman, C.S., 2012. Framing the issue of research quality in a context of research diversity. Account. Horiz. 26, 821–831.
Chenhall, R.H., Moers, F., 2007. The issue of endogeneity within theory-based, quantitative management accounting research. Eur. Account. Rev. 16, 173–195.
Chenhall, R.H., Smith, D., 2011. A review of Australian management accounting research: 1980–2009. Account. Finance 51, 173–206.
Chow, C.W., 2010. Looking back, peeking ahead. J. Manage. Account. Res. 22, 67–72.
Coloquitt, J.A., Zapata-Phelan, C.P., 2007. Trends in theory building and theory testing: a five-decade study of the Academy of Management Journal. Acad. Manage. J. 50, 1281–1303.
Covaleski, M.A., Evans III, J.H., Luft, J.L., Shields, M.D., 2003. Budgeting research: three theoretical perspectives and criteria for selective integration. J. Manage. Account. Res. 15, 3–49.
Dahlstrom, R., Nygaard, A., Crosno, J.L., 2008. Strategic, metric, and methodological trends in marketing research and their implications for future theory and practice. J. Market. Theor. Pract. 16, 139–152.

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

Davilla, T., Oyon, D., 2008. Cross-paradigm collaboration and the advancement of management accounting knowledge. Crit. Perspect. Account. 19, 887–893.
DeFond, M.L., 2015. Annual report and editorial commentary for The Accounting Review. Account. Rev. 90, 2603–2638.
Demski, J.S., 2007. Is accounting an academic discipline? Account. Horiz. 21, 153–157.
Drost, E.A., 2011. Validity and reliability in social science research. Educ. Res. Perspect. 38, 105–123.
Drymiotes, G.C., 2011. Information precision and manipulation incentives. J. Manage. Account. Res. 23, 231–258.
Edwards, J.R., Dean, G., Clarke, F., 2013. Accounting academic elites: the tale of ARIA. accounting. Organ. Soc. 38, 365–381.
Evans III, J.H., Feng, M., Hoffman, V.B., Moser, D.V., Van der Stede, W.A., 2015. Points to consider when self-assessing your empirical accounting research. Contemp. Account. Res. 32, 1162–1192.
Ferreira, L.D., Merchant, K.A., 1992. Field research in management accounting and control: a review and evaluation. Account. Audit. Account. J. 5, 3–34.
Foster, G., Young, S.M., 1997. Frontiers of management accounting research. J. Manage. Account. Res. 9, 63–77.
Gioia, D.A., Pitre, E., 1990. Multiparadigm perspectives on theory building. Acad. Manage. Rev. 15, 584–602.
Gray, R., Milne, M.J., 2015. It's not what you do, it's the way you do it? Of method and madness. Crit. Perspect. Account. 32, 51–66.
Hamann, P.M., Schiemann, F., Bellora, L., Guenther, T.W., 2013. Exploring the dimensions of organizational performance: a construct validity study. Organ. Res. Methods 16, 67–87.
Harrison, D.A., Klein, K.J., 2007. What's the difference? Diversity constructs as separation, variety, or disparity in organizations. Acad. Manage. Rev. 32, 1199–1228.
Hartmann, F.G.H., Moers, F., 1999. Testing contingency hypotheses in budgetary research: an evaluation of the use of moderated regression analysis. Account. Organ. Soc. 24, 291–315.
Hermanson, D.R., 2015. Model 2-A personal journey in pursuit of creativity and impact. Crit. Perspect. Account. 26, 130–140.
Hesford, J.W., Lee, S.-H., Van der Stede, W.A., Young, S.M., 2007. Management accounting: a bibliographic study. In: Chapman, C.S., Hopwood, A.G., Shields, M.D. (Eds.), Handbook of Management Accounting Research, vol. 1. Elsevier, Amsterdam et al, pp. 3–26.
Homburg, C., Klarmann, M., Reimann, M., Schilke, O., 2012. What drives key informant accuracy? J. Market. Res. 49, 594–608.
Hopper, T., Bui, B., 2016. Has management accounting research been critical? Manage. Account. Res. 31, 10–30.
Hopper, T., Otley, D., Scapens, B., 2001. British management accounting research: whence and whither: opinions and recollections. Br. Account. Rev. 33, 263–291.
Hopwood, A.G., 2007. Whiter accounting research? Account. Rev. 82, 1365–1374.
Hopwood, A.G., 2008a. Changing pressures on the research process: on trying to research in an age when curiosity is not enough. Eur. Account. Rev. 17, 87–96.
Hopwood, A.G., 2008b. Management accounting research in a changing world. J. Manage. Account. Res. 20, 3–13.
Hoque, Z., Covaleski, M.A., Gooneratne, T.N., 2013. Theoretical triangulation and pluralism in research methods in organizational and accounting research. Account. Audit. Account. J. 26, 1170–1198.
Humphrey, C., Gendron, Y., 2015. What is going on? The sustainability of accounting academia. Crit. Perspect. Account. 26, 47–66.
Ittner, C.D., Larcker, D.F., 2001. Assessing empirical research in managerial accounting: a value-based management perspective. J. Account. Econ. 32, 349–410.
Ittner, C.D., Larcker, D.F., 2002. Empirical managerial accounting research: are we just describing management consulting practice? Eur. Account. Rev. 11, 787–794.
Ittner, C.D., 2014. Strengthening causal inferences in positivist field studies. Account. Organ. Soc. 39, 545–549.
Jacobs, K., Marcon, G., Witt, D., 2004. Cost and performance information for doctors: an international comparison. Manage. Account. Res. 15, 337–354.
Kaplan, R.S., 1986. The role for empirical research in management accounting. Account. Organ. Soc. 11, 429–452.
Kaplan, R.S., 1994. Management accounting (1984–1994): Development of new practice and theory. Manage. Account. Res. 5, 247–260.
Kaplan, R.S., 2006. The competitive advantage of management accounting. J. Manage. Account. Res. 18, 127–135.
Kaplan, R.S., 2011. Accounting scholarship that advances professional knowledge and practice. Account. Rev. 86, 367–383.
Keating, P.J., 1995. A framework for classifying and evaluating the theoretical contributions of case research in management accounting. J. Manage. Account. Res. 7, 66–86.
Klemstine, C.F., Maher, M.W., 1984. Management Accounting Research: A Review and Annotated Bibliography. Routledge, New York and London.
Krishnan, R., 2015. Management accountant – what ails thee? J. Manage. Account. Res. 27, 177–191.
Kumar, N., Stern, L.W., Anderson, J.C., 1993. Conducting interorganizational research using key informants. Acad. Manage. J. 36, 1633–1651.
Langfield-Smith, K., 2008. Strategic management accounting: how far have we come in 25 years? Accounting. Audit. Account. J. 21, 204–228.
Lillis, A.M., Mundy, J., 2005. Cross-sectional field studies in management accounting research – closing the gaps between surveys and case studies. J. Manage. Account. Res. 17, 119–141.
Lillis, A.M., 1999. A framework for the analysis of interview data from multiple field research sites. Account. Finance 39, 79–105.
Lindquist, T.M., Smith, G., 2009. Content and citation analysis of the first 20 years. J. Manage. Account. Res. 21, 249–292.
Llewellyn, S., Northcott, D., 2005. The average hospital. Account. Organ. Soc. 30, 555–583.
Luft, J., Shields, M.D., 2002. Zimmerman's contentious conjectures: describing the present and prescribing the future of empirical management accounting research. Eur. Account. Rev. 11, 795–803.
Luft, J., Shields, M.D., 2003. Mapping management accounting: graphics and guidelines for theory-consistent empirical research. Account. Organ. Soc. 28, 169–249.
Luft, J., Shields, M.D., 2014. Subjectivity in developing and validating causal explanations in positivist accounting research. Account. Organ. Soc. 39, 550–558.
Lukka, K., Granlund, M., 2002. The fragmented communication structure within the accounting academia: the case of activity-based costing research genres. Account. Organ. Soc. 27, 165–190.
Lukka, K., Kasanen, E., 1995. The problem of generalizability: anecdotes and evidence in accounting research. Accounting. Audit. Account. J. 8, 71–90.
Lukka, K., Mouritsen, J., 2002. Homogeneity or heterogeneity of research in management accounting? Eur. Account. Rev. 11, 805–811.
Lukka, K., 2010. The roles and effects of paradigms in accounting research. Manage. Account. Res. 21, 110–115.
Maher, M.W., 2001. The evolution of management accounting research in the United States. Br. Account. Rev. 33, 293–305.
Maines, L.A., Salamon, G.L., Sprinkle, G.B., 2006. An information economic perspective on experimental research in accounting. Behav. Res. Account. 18, 85–102.
Malmi, T., Granlund, M., 2009. In search of management accounting theory. Eur. Account. Rev. 18, 597–620.
Merchant, K.A., Van der Stede, W.A., 2006. Field-based research in accounting: accomplishments and prospects. Behav. Res. Account. 18, 117–134.
Merchant, K.A., Van der Stede, W.A., Zheng, L., 2003. Disciplinary constraints on the advancement of knowledge: the case of organizational incentive systems. Account. Organ. Soc. 28, 251–286.
Merchant, K.A., 2010. Paradigms in accounting research: a view from North America. Manage. Account. Res. 21, 116–120.
Merchant, K.A., 2013. Looking back, looking forward. J. Manage. Account. Res. 25, 65–70.
Messner, M., 2015. Research orientation without regrets. Crit. Perspect. Account. 26, 76–83.
Mitchell, F., 2002. Research and practice in management accounting: improving integration and communication. Eur. Account. Rev. 11, 277–289.
Mittendorf, B., 2015. Management accounting research: knowledge base, themes, and future directions. J. Manage. Account. Res. 27, 121–122.
Modell, S., 2005. Triangulation between case study and survey methods in management accounting research: an assessment of validity implications. Manage. Account. Res. 16, 231–254.
Modell, S., 2009. In defence of triangulation: a critical realist approach to mixed methods research in management accounting. Manage. Account. Res. 20, 208–221.
Modell, S., 2014. The societal relevance of management accounting: an introduction to the special issue. Account. Bus. Res. 44, 83–103.
Moers, F., 2007. Doing archival research in management accounting. In: Chapman, C.S., Hopwood, A.G., Shields, M.D. (Eds.), Handbook of Management Accounting Research, vol. 1. Elsevier, Amsterdam et al, pp. 399–413.
Moser, D.V., 2012. Is accounting research stagnant? Account. Horiz. 26, 845–850.
Nikolaev, V., van Lent, L., 2005. The endogeneity bias in the relation between cost-of-debt capital and corporate disclosure policy. Eur. Account. Rev. 14, 677–724.
Oler, D.K., Oler, M.J., Skousen, C.J., 2010. Characterizing accounting research. Account. Horiz. 24, 635–670.
Otley, D., 2003. Management control and performance management: whence and whither? Br. Account. Rev. 35, 309–326.
Raffournier, B., Schatt, A., 2010. Is European accounting research fairly reflected in academic journals? An investigation of possible non-mainstream and language barrier biases. Eur. Account. Rev. 19, 161–190.
Salterio, S.E., 2015. Barriers to knowledge creation in management accounting research. J. Manage. Account. Res. 27, 151–170.
Scandura, T.A., Williams, E.A., 2000. Research methodology in management: current practices, trends, and implications for future research. Acad. Manage. J. 43, 1248–1264.
Scapens, R.W., Bromwich, M., 2001. Management accounting research: the first decade. Manage. Account. Res. 12, 245–254.
Scapens, R.W., Bromwich, M., 2010a. Practice, theory and paradigms. Manage. Account. Res. 21, 77–78.
Scapens, R.W., Bromwich, M., 2010b. Management accounting research: 20 years on. Manage. Account. Res. 21, 278–284.
Scapens, R.W., 2006. Understanding management accounting practices: a personal journey. Br. Account. Rev. 38, 1–30.

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002

Shadish, W.R., Cook, T.D., Campbell, D.T., 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference, second ed. Houghton, Boston and New York.
Shields, M.D., 1997. Research in management accounting by North Americans in the 1990. J. Manage. Account. Res. 9, 3–61.
Shields, M.D., 2015. Established management accounting knowledge. J. Manage. Account. Res. 27, 123–132.
Snow, C.C., Thomas, J.B., 1994. Field research methods in strategic management: contributions to theory building and testing. J. Manage. Stud. 31, 457–480.
Soin, K., Collier, P., 2013. Risk and risk management in management accounting and control. Manage. Account. Res. 24, 82–87.
Tucker, B., Lowe, A.D., 2014. Practitioners are from Mars, academics from Venus? An investigation of the research-practice gap in management accounting. Account. Audit. Account. J. 27, 394–425.
Tucker, B., Parker, L., 2014. In our ivory towers? The research-practice gap in management accounting. Account. Bus. Res. 44, 104–143.
Unerman, J., Chapman, C.S., 2014. Academic contributions to enhancing accounting for sustainable development. Account. Organ. Soc. 39, 385–394.
Van Campenhout, G., Van Caneghem, T., 2010. Article contribution and subsequent citation rates: evidence from European Accounting Review. Eur. Account. Rev. 19, 837–855.
Van der Stede, W.A., Young, S.M., Chen, C.X., 2005. Assessing the quality of evidence in empirical management accounting research: the case of survey studies. Accounting. Organi. Soc. 30, 655–684.
Van der Stede, W.A., 2011. Management accounting research in the wake of the crisis: some reflections. Eur. Account. Rev. 20, 605–623.
Van der Stede, W.A., 2014. A manipulationist view of causality in cross-sectional survey research. Account. Organ. Soc. 39, 567–574.
Van der Stede, W.A., 2015. Management accounting: where from, where now, where to? J. Manage. Account. Res. 27, 171–176.
Wagenhofer, A., 2016. Exploiting regulatory changes for research in management accounting. Manage. Account. Res. 31, 112–117.
Wooldridge, J.M., 2013. Introductory Econometrics – A Modern Approach, fifth ed. Cengage, Learning Boston (MA).
Yin, R.K., 2014. Case Study Research: Design and Methods, fifth ed. Sage, Thousand Oaks (CA) et al.
ter Bogt, H.J., Scapens, R.W., 2012. Performance management in universities: effects of the transition to more quantitative measurement systems. Eur. Account. Rev. 21, 451–497.
Zimmerman, J.L., 2001. Conjectures regarding empirical management accounting research. J. Account. Econ. 32, 411–427.

Please cite this article in press as: Lachmann, M., et al., Diversity and validity in positivist management accounting research—A longitudinal perspective over four decades. Manage. Account. Res. (2016), http://dx.doi.org/10.1016/j.mar.2016.07.002