

Performance Measurement at Universities: studying Function and Effect of Student Evaluations of Teaching

KLARISSA LUEG

Aarhus University, Denmark

This paper proposes empirical approaches to testing the reliability, validity, and organizational effectiveness of student evaluations of teaching (SET) as a performance measurement instrument in knowledge management at the institutional level of universities. Departing from Weber's concept of bureaucracy and critical responses to this concept, we discuss how contemporary SET are used as an instrument of organizational control at Danish universities. A discussion of the current state of performance measurement within the frame of new public management (NPM) and its impact on knowledge creation and legitimation forms the basis for proposing four steps of investigation. The suggested mixed-methods approach comprises the following: first, thematic analysis can serve as a tool to evaluate the legitimacy discourse as initiated by official SET affirmative documents by government, university, and students. Second, constructs for the SET questionnaire can be developed and compared to existing SET questionnaires in terms of reliability and validity. Third, data from SET can be used to corroborate the relationship between the qualitative (comments) and quantitative (scaled questionnaire) sections. Fourth, it can be investigated if SET actually contribute to teaching improvement by examining how the instrument is integrated into systematic ex-ante and ex-post organizational management. It is expected to find discrepancy between the proponents' intent to evaluate teaching and the way the performance measurement instrument is implemented.

Key words: student evaluations of teaching; performance measurement; performance management systems; higher education; knowledge management; knowledge economy

1. INTRODUCTION

“Often, we get so wrapped up in the measuring of performance that we forget to examine the purposes for which we measure.” (Bromberg, 2009: 214)

During the last decade, at the latest, many European universities had to undergo rapid changes from their traditional administrative forms to new public management. Politicians and other advocates proffered efficiency improvements and accountability of a hitherto seemingly intransparent organization. Even though one common and consistent definition of new public management (NPM) is missing and it can merely be considered an “umbrella term” (Van de Walle & Hammerschmid, 2011: 191), the notion usually implies the transfer of free market economy practices to public organizations, such as universities. The doctrinal compounds of NPM have first been summarized by Hood (1991) and consist of, shortened, a “hands-on professional management”, performance standards and measurements, output control, the disaggregation of units, competition, private sector management style, and cost-efficient resource use (Hood, 1991: 5). This reinvention of the university organization widely across Europe (Schubert, 2009; Van de Walle & Hammerschmid, 2011) has in parts been presented as a logical solution to university administration (Choong, 2013), but has also – to a larger extent – led to strong criticism (Andersen, 2002; Clark, 1998; Evans, 2004; Kallio & Kallio, 2012; Moed, Burger, Frankfort, & van Raan, 1985; Stölting & Schimank, 2001; Temple, 2014; Ward, 2011). These new rules that were enforced on universities and researchers have changed the power balance in the organization, the position of the organization in society, and the essential understanding of “knowledge”. Struggles relate to the question of the legitimacy of knowledge (Lyotard, 1984: 6, s. also Temple 2014): Who, within an economized system, decides what knowledge is and who knows what needs to be decided?” (Lyotard, 1984: 9). In the focus of criticism were also the doctrine of performance measurement systems and instruments (PMI) that were applied to researcher and teacher performance (Kallio & Kallio, 2012). In Denmark, this change has been implemented in 2003 as “the new 2003 University Act” (Kristensen, Nørreklit, & Raffnsøe-Møller, 2011; Wright & Williams Ørberg, 2008) and this implementation almost reinvented the university as a research and teaching organization. In their extended review of the changes, Kristensen et al. (Kristensen et al., 2011, s. also Andersen 2002) list a shift in the management structure to mostly external agents and the connection of resource allocation to measurements of research output, but, interestingly enough, not to university administration (Kristensen et al., 2011: 14).

This new approach to university management conceptualized universities as private, competitive and output-oriented businesses and brought substantial change to how the university structure was managed, how resources were allocated and to how performance was managed (Kristensen et al., 2011). Despite pointing out that the university is “probably the most ‘performance-measuring’ institution there is” (Raffnsøe-Møller, 2011: 49), criticism amongst Danish scholars was huge. Researchers argue that despite the official Danish government agenda to liberate the universities from state influence, the universities are still under pressure from the state – namely to live up to economized standards (Wright & Williams Ørberg, 2008).

This paper focusses on the concept of student teacher evaluations (SET) as performance measurement instrument. Departing from Max Weber's and Robert K. Merton's concepts of bureaucracy and the perspective of new institutionalism (DiMaggio & Powell, 1983) it deals with how such a PMI can be investigated from a sociological perspective. The research site of this study are the Danish universities, amongst them Aarhus university (AU). The topic of university performance measurement is highly relevant for European governments who are accountable to their citizens for the quality of public universities; to students who aim at obtaining a high quality education; for teachers (researchers) in HE whose careers depend on SET and for both agent groups whose working milieu will be affected by SET (Kallio & Kallio, 2012). For these very reasons, if SET have to be installed at universities, their flawless, and thus fair, functioning is a necessity. To approach this issue empirically, I propose a four part study, that tests the reliability, validity, and organizational effectiveness of student teaching evaluations (SET) as a performance measurement instrument in universities. Further, I aim to provide theory-guided constructs for future SET development. The novelty of this project lies in its character as a theory-driven mixed-methods empirical test of a selected, highly relevant PMI which will contribute to the improvement of university performance measurement systems. This paper is thus to be understood as an outline of a larger future research project and a discussion of the status quo of SET. The remaining part of this paper proceeds as follows: Section 1 is concerned with background information on performance measurement and SET as customer-client relation. The second section introduces the theoretical perspective on the study. Section 3 derives a couple of research questions, followed by an outline of a possible methodology to pursue each of these questions (section 4). Finally, I discuss this project's implications, its limitations and provide a brief recapitulatory conclusion in section 5.

2. PERFORMANCE MEASUREMENT AND STUDENT TEACHER EVALUATION AS CUSTOMER-CLIENT COMMUNICATION

The term performance measurement originates from the field of management accounting and comprises a merely quantitative approach as well as a broader approach to measurement. Neely et al.'s definition exemplifies the quantitative approach by defining performance measurement as "the process of quantifying the efficiency and effectiveness of past actions." (Neely, Adams, & Kennerley, 2002: xiii). A broader approach, which includes qualitative measures, and the one that this study will base its design on, is the one provided by Moullin: performance measurement means "evaluating how well organisations are managed and the value they deliver for customers and other stakeholders" (Moullin, 2007: 188). Since new public management seems to construct performance measurement systems as measuring university "excellence", I adopt Moullin's suggestion to link the definitions of excellence and performance measurement. "Organisational excellence is outstanding practice in managing organisations and delivering value for customers and other stakeholders" (Moullin, 2007: 182). I am aware that these definitions can be problematized since they seem to capture the features of a leadership tool rather than provide

definitions from a critical meta-perspective and thus subordinate to organizational hegemony: the character of 'how well' processes work, 'efficiency', 'excellence' or basically 'performance' itself (s. also Otley, 1999: 364) seems to comply with the leadership perspective. Yet, more critical perspectives from sociological or educational studies have not provided competing definitions. Agents that are given the opportunity to use performance measurements instruments as the evaluators of another agents' performance will find themselves in a position of control. In this case, this control is given to university students from their first semester on. Course evaluations at AU usually offer evaluation with a mix of Likert-style scales and other rating scales and inquire about the students' satisfaction with the class framework – e.g. the physical settings – but also about knowledge-related aspects. Knowledge related aspects are e.g. "I find the objectives of the course appropriate" or "The ability of the teacher to put the topic and the solutions to the assignment into perspective" or "The contribution of the lectures/classes to your knowledge and comprehension" (AU, 2013). In Denmark, course evaluations are usually planned to be summative and completely anonymous. They are usually distributed to the students in the form of a standardized questionnaire, with some room for additional comments.

This strengthening of the role of the student as a stakeholder or even as a customer (Kristensen et al., 2007), has additionally heated the debate about new public management and the ethics behind performance measurement. While proponents argue that SET lead to increased teaching and learning quality, critics suspect negative impacts on staff motivation and compliant behavior as a self-protective measure (Crumbly & Reichelt, 2009; Simpson & Sigauw, 2000), misinterpretation, abuse, even cruelty by students (Chan, Luk, & Zeng, 2014; Clayson, 2005; Hajdin & Pažur, 2012; Lindahl & Unger, 2010) or simply a lack of validity in indicating student achievements (Galbraith, Merrill, & Kline, 2012). For the aspect of quality knowledge mediation the aspect of self-protection by the teacher (researcher) and compliant behavior is crucial: if the SET are used as criteria for tenure or promotions, teachers (researchers) behave rationally by decreasing the course level to the lowest common denominator in order to receive favorable SET. Thus, content and quality of knowledge mediation, even the basic understanding of what knowledge is, are affected by SET. Not the initial expert in a field of knowledge – the class instructor – has interpretation authority, but the request to let students assess aspects such as "The ability of the teacher to put the topic [...] into perspective" strongly suggests a very high insight and previous knowledge by the students. Some SET, though not at AU, go so far as to ask "Is your teacher knowledgeable?" (Platt, 1993). Knowledge is constructed as a product that the customer has the right to assess; the researcher is despite the strengthened demand for research-based teaching not created as the producer, but reduced to the role of a shop assistant. It can be argued that this idea of a teaching researcher has a corrupting effect on the students' attitudes toward higher education and their understanding of their own role (Platt, 1993). An informed customer, so suggests the standardized questionnaire, may equally judge the shop, the shop assistant and the product. Rational, self-protective behavior will lead class instructors to adjust their class and what they present as intended learning outcomes respectively knowledge in a way that also the poorer, unprepared students can follow. If this behavior is seemingly justified by evaluations, the course will most likely be repeated with the same intended learning outcomes, and thus, a new knowledge standard is legitimized. These outlined negative consequences are appropriate arguments to make a case against SET in its very

essence. Proposing an alternative approach, this project aims to ask whether SET have a (positive) impact at all or if they are an empty end-in-itself control instrument. In short, do SET really function as an instrument of accountability? Thus, this paper does not condone downright criticism towards university performance measurement and SET and the demand to abolish them. Essentially though, I argue that proponents do not sufficiently consider the reliability and validity of their measures and the function and effect of the instrument in general. The question whether universities have developed SET with “meaningful measures that drive performance improvement” (Moullin, 2007: 181) is highly relevant.

Isolated instruments like SET can only be considered reliable and valid when embedded in a comprehensive performance measurement system which motivates agents and defines what performance is and to whom it has to be delivered. Failing to account for the special conditions and agent motivation in knowledge management may result in a decrease in teaching quality, functional fixation on single measures, and alienation from knowledge mediation. Dysfunctional consequences instead of performance improvement or control are possible. Thus, to be functional, SET have to be embedded in a mature management context: close communication with the lecturers on meaning, measurement and operation of the SET and the execution of effective organizational consequences and practices depending on the results – thus the meaning of control – are just two examples of such embedment. These ex-ante and ex-post steps are taken to foreclose that an illusion of control is created, where really no control is present (Rosanas & Velilla, 2005). Meaning and purpose, in particular, have to be defined, made transparent and adapted to organizational changes: “Any controlled system requires objectives and goals against which its performance can be assessed” (Otley, 1999: 367). On a more operational level, the student questionnaires must be designed to prevent bias. Misconstruction of any part of a performance measurement system may otherwise lead to a negative work environment and to “incorrect inferences in decision-making” (Choong, 2013: 102). Bromberg (2009: 214) points out that the lack of “sophisticated measures” is the highest threat to the purpose “of improved productivity”. This study will enable identification of how the SET instrument should function departing from our chosen theoretical approaches, how it is meant to function by its initiators and to test if the chosen operationalization and contextualization fulfill these intentions. Eventually, results may enable researchers and university management to propose or define avenues for organizational improvement.

3. THEORETICAL PERSPECTIVES

From a theoretical standpoint, this project aims at making three theoretical propositions meet. The perspective combines the Weberian sketch of the ideal bureaucracy (Weber, 1905, 1968) with Robert K. Merton’s critical approach towards bureaucracy’s dysfunctions (Merton, 1940), and the perspectives of organization-related critical management studies (Kanter, 1977; Parker, 2002) as well as neo-institutionalism and isomorphism (DiMaggio & Powell, 1983). Max Weber’s concept of the shell as hard as steel serves as an overarching theoretical construct. Ward’s recent work on the “global restructuring of knowledge and education” (Ward, 2011) has been proven helpful in applying such concepts to concrete HE practices of control and government. Further, I will find addenda to Weber’s approach on bureaucracy and to the perspective of critical management to

authority and power in Pierre Bourdieu's concept of symbolic power (Bourdieu, 1979, 1991). I alert that the sociological framework will provide a critical perspective, whilst, for the purpose of the empirical test, I will take on the stance and the notions of accounting tradition, thus in principle affirming organizational control.

3.1. Bureaucracy – Max Weber

The core of this project will be approached from the perspective of Max Weber's theory of bureaucracy. This choice of Weber's critical, yet in some principles condoning theoretical approach towards administration is a deliberate one: Like Weber, this project assumes that bureaucracy is (can be) the most efficient way to organize human work and outcome, but I point to the danger of its paralyzing effect when not being thoroughly designed. This paralyzing effect is what leads Weber to his notion of the "shell as hard as steel" ("stahlhartes Gehäuse", Weber, 1905) mostly known as "iron cage" as translated by Talcott Parsons (for a discussion of notion and translation s. Baehr, 2001). A powerful bureaucracy controls society's individuals – but who controls the bureaucracy? An uncontrolled bureaucracy threatens to limit individual freedom and human life can be trapped in a shell of rule-based control. Despite this criticism, Weber defined rules about an ideally working bureaucracy: it is characterized by a hierarchical organization, exact assignments of competences, selection of staff and promotion based on competence as well as seniority (judged by the organization, not individuals), documentation and legitimation of all processes in writing (transparency), application of consistent rules and regulations, expert training for bureaucratic officials, and the fact that rules are implemented by neutral officials (separation of ownership and control) (Allan, 2009; Hartfiel, 1976). Bureaucracy, in short, should ideally protect individuals against meaningless orders and conducts as well as abusive domination.

3.2. Dysfunctions of bureaucracy – Robert K. Merton

Robert K. Merton's reflections on the negative sides and the dangers of bureaucracy can, to a certain extent, be understood as an answer to Weber's overarching theory: e.g. whilst Weber still defines bureaucracy as "fundamentally domination through knowledge" (Weber, 1968: 225), Merton casts doubt on the "knowledge" determinant by emphasizing the agency of bureaucrats and flaws due to inflexibility and lack of adapting skills and procedures to varying circumstances (Merton, 1957). Merton's outlook on overly conform behavior and the categorical use of rules stresses the risk of a bureaucracy becoming a system that controls for controlling purposes, and for this very reason, may remain ineffective. In the worst case, the bureaucracy's and bureaucrats' "conformity to the rules interferes with the achievement of the purposes of the organization, in which case we have the familiar phenomenon of the technicism or red tape of the official. An extreme product of this process of displacement of goals is the bureaucratic virtuoso, who never forgets a single rule binding his action and hence is unable to assist many of his clients." (Merton, 1940: 563).

3.3. Neo-Institutionalism/Isomorphism

This perspective, though Merton is not amongst the most commonly used theorists for Critical Management Studies (CMS), is mirrored in modern and critical management studies: focused on the meso-level of business organizations, modern and CMS provide possible connections to these thoughts: unnecessary or unsuitable bureaucracy must be eliminated in order for an organization to work efficiently (Wren & Bedeian, 2009). Critical management studies emphasize the reproduction of power structures, social injustice power rituals by administration and management in organizations. CMS often overlap with the approach of neo-institutionalism which attempts to understand how organizational developments affect the behavior of agents. The most relevant trend within neo-institutionalism for this study is research on coercive isomorphism: organizations, though they might be different in their core, appear to be managed and structured in a similar way due to political influences and legitimacy struggles. This also enables bad practices to diffuse (Strang & Macy, 2001). Neo-institutionalism acknowledges the pressure on organizations to strive for legitimacy in the broader environment of institutions. Public institutions are also being rationalized, mimicking the processes of private business (DiMaggio & Powell, 1983) – whilst ignoring their very different roots and different identity concepts. Partly supporting Merton’s criticism – though Merton focuses more on the actor – DiMaggio and Powell claim that the similarization of bureaucracies across organizations does not necessarily make them “more efficient” (DiMaggio & Powell, 1983: 147). Since DiMaggio and Powell, many researchers have critically remarked that the homogenization of organizations and their bureaucratic structures has reached the university sector. Using the example of SET this project sets out to examine whether this is to the university’s benefit, detriment or if it has any effect at all.

4. FUTURE RESEARCH: QUESTIONS TO CONSIDER

This paper essentially departs from the assumptions that a) effective measurement of teacher performance is possible and that b) the current measurements are insufficient and erroneous. To control these assumptions, I propose that future research considers five interrelated research questions. I describe and motivate these research questions in the following.

It can be assumed that politicians, university administration and leadership tie goals to SET that go beyond plain performance measurement as an end in itself, but aim at performance improvement or ensuring good quality. Knowing the stated purpose of the SET and what is defined as good or desirable performance will build the foundation for the following research questions. Thus, it has to be asked

1. What are the intended purposes of SET?

The underlying broad definition of performance measurement systems (s. earlier) allows for a variety of measuring methods, including qualitative approaches. For this reason and to make it possible to develop a coherent basis for possible recommendations on improving SET later on, it has to be asked

2. Is the instrument appropriate? [Are there alternative possible measurement instruments to serve these purposes?]

Once the intended purposes and the appropriateness of the choice of SET as a measuring instrument are known a shift to the operational level is meaningful. Departing from the findings of what SET are supposed to measure a comprehensive set of questionnaire constructs can be developed (Hair, Black, Babin, & Anderson, 2006). Actual SET from Danish universities should be used and their shortcomings can be investigated compared to the theory-driven development of SET constructs. On this basis, it has to be asked

3. Is the questionnaire valid? [Do the questions load positive on the underlying constructs and do they measure what they are intended to measure?]

Knowing the purposes of the PMI creators and proponents, it is now possible to focus on the actors assessing the lecturers' performance: in order for the questionnaire design to be a meaningful instrument, the students' perception of its purposes has to be similar to the intended ones. A standardized questionnaire can only be *relevant* if all possible categories of student perceptions are considered and taken into account. Ignorant questionnaire design will lead to biased answering behavior, which is time-consuming to detect and significantly decreases the value of the instrument (Choi & Pak, 2005). Surveyed individuals might tend to indicate their discontent on a Likert scale even if the question does not relate to the cause of their unhappiness (extreme responding, negativity bias). Considering that a student body can be highly diverse (age, gender, study level, learning motivation, programs, cultural expectations etc.) it can be doubted that these students can be surveyed using the same questionnaire. Thus, it has to be asked:

4. Is the questionnaire relevant? [Does the instrument sufficiently consider different student types and groups and their expectations towards teaching?]

Effective PMIs require systematic performance management of processes that will support reaching the intended purposes. This could for instance include follow-up communication with the teaching staff, incentives and offers for teaching improvement where needed, or rewards for high quality teaching (s. earlier). It is thus crucial to ask:

5. Is the instrument effective? [What organizational and managerial steps, involving the teaching staff, are taken after the questionnaires are statistically processed?]

5. OUTLINE OF A POSSIBLE METHODOLOGY

The outlined research questions can be handled by employing a mixed-methods approach for the four steps of investigation.

First, thematic discourse analysis (Fereday & Muir-Cochrane, 2006; Guest, MacQueen, & Namey, 2012) can be used to evaluate documents focusing on the support of SET. These documents could originate from the three agencies government, university and student representatives. This qualitative approach to content analysis is chosen to ensure that all categories, including non-political ones, are made visible by a first analysis. Thereby, the reasoning of these three players for supporting SET and – since SET are already an established instrument at all Danish universities – the legitimacy discourse around measuring and assessing knowledge built by the three agencies can be described. In addition, this will enable the researcher to trace if the discourse is referential, that is, to what extent and with what weighting these three agencies have cross reference to each other. The result of this investigation will be an overview of the purpose that SET should ideally serve and of what they are intended to measure. This will answer the research questions 1. *What are the intended purposes of SET?* and 2. *Is the instrument appropriate?*

As a *second*, consecutive step, guiding constructs for the SET questionnaire design are to be remodeled and re-derived. With the help of these constructs questionnaires to contrast with the existing SET questionnaires can be developed (Dillman, 2007; Hair et al., 2006; Luft & Shields, 2003). This will enable the researcher to test the questionnaires for their reliability and validity in cases where in-classroom surveying is being hindered by the already existing SET, the confusion this might cause amongst students and the unfavorable dependence on real-time class semester schedules. Equally, in this step erroneous or misleading questionnaire design can be identified (such as using a Likert-similar scale both for attitudes and keywords or sentence fragments, (AU, 2013)). Overlaps as well as shortcomings of the supporters' intent to evaluate teaching and the way the PMI is applied can be identified. Thus, research question 3. *Is the questionnaire valid?* can be answered.

Thirdly, the relationship between qualitative (comment section) and quantitative (scaled questionnaire) information can be tested. Since practice shows that usually only the quantitative measures are taken into account, identify student clusters based on the qualitative section (e.g., overburdened vs. under-challenged; content-focused vs. holistic) should be identified as a first step. This information can then be corroborated with the scaled constructs. It is expected that strongly diverging interpretations of the same grade given in an evaluation can be found. Consequently, research question 4. *Is the questionnaire relevant?* can be answered.

The fourth step investigates the actual consequences of the PMI compared to its proposed effect (s. thematic analysis) of teaching improvement. This ex-post observation focuses on administrative and management measurements taken after the assessment of the evaluation data. It should be traced what happens after the SET are statistically processed. This means that official policies on the follow-up-process from communicating the results to the teachers to reactions towards the quality of teaching (e.g., rewards and incentives, improvement initiatives, monitoring

of change) should be identified. The concrete method of investigation depends on the tacitness of information: Preferably, written documents should be analyzed, but where departmental and university policy on SET follow up initiatives is not textually documented, (qualitative) interviews with the decision-makers in question should be considered. Additionally, a teacher-typology based on a survey across two departments will demonstrate the effect of SET on a personal level. This will allow for answering question 5. *Is the instrument effective?*

6. DISCUSSION

This paper introduced an outline of possible empirical approaches to measuring the effectiveness and function of student evaluations of teaching. It has taken a critical stance on SET as functioning performance measurement instrument, pronouncing its negative impact on the perception and legitimation of knowledge, whilst generally condoning performance measurement of higher education staff. The outlined and interrelated subprojects built a bottom-up approach that fosters an understanding of the intended purposes of SET and then, adjusting to this perspective, reconstructs whether these purposes can be served. Future results will have to be interpreted within the frame of the Danish context, since other countries might have different SET traditions and a different understanding of the student-teacher relationship.

The findings might support governments and universities in successfully introducing SET while ensuring an organization-specific approach and reliability and validity of the instrument. An appropriate PMI will do justice to a diverse student body by de facto including their patterns of thought. Further, insights will hopefully inspire successful, bottom-up inspired *management* of staff performance, thereby ensuring the effectiveness of SET and at the same time allowing for an affirmative and motivating performance measurement system. This extension of a hitherto single performance measurement *instrument* to a management system will allow for addressing the problem of legitimizing complied knowledge and lowered course content standard: teachers whose performance is embedded in a sensitive and appropriate set of communication and incentives will behave rational when improving their course content instead of lowering their expectations.

REFERENCES

Allan, K. D. 2009. *Explorations in Classical Sociological Theory: Seeing the Social World*. Thousand Oaks, CA: Sage.

Andersen, H. 2002. Universitetsreformen: Topstyring og ensretning. *Dansk Sociologi*, 13(4): 80-88.

AU. 2013. Course Evaluation. Aarhus: Aarhus University.

Baehr, P. 2001. The "Iron Cage" and the "Shell as Hard as Steel": Parsons, Weber, and the Stahlhartes Gehäuse Metaphor in the Protestant Ethic and the Spirit of Capitalism. *History and Theory*, 40(2): 153-169.

- Bourdieu, P. 1979. Symbolic Power. *Critique of Anthropology*, 4(13-14): 77-85.
- Bourdieu, P. 1991. *Language and Symbolic Power*. Cambridge, MA: Harvard University Press.
- Bromberg, D. 2009. Performance Measurement: A System with a Purpose or a Purposeless System? *Public Performance & Management Review*, 33(2): 214-221.
- Chan, C. K. Y., Luk, L. Y. Y., & Zeng, M. 2014. Teachers' perceptions of student evaluations of teaching. *Educational Research and Evaluation*, 20(4): 275-289.
- Choi, B. C. K., & Pak, A. W. P. 2005. A catalog of biases in questionnaires. *Preventing Chronic Disease*, 2(1).
- Choong, K. K. 2013. Understanding the features of performance measurement system: a literature review. *Measuring Business Excellence*, 17(4): 102-121.
- Clark, B. R. 1998. *Creating entrepreneurial universities: organizational pathways of transformations*. New York: Elsevier.
- Clayson, D. E. 2005. Within-Class Variability in Student-Teacher Evaluations: Examples and Problems. *Decision Sciences Journal of Innovative Education*, 3(1): 109-124.
- Crumbly, L. D., & Reichelt, K. J. 2009. Teaching effectiveness, impression management, and dysfunctional behavior. *Quality Assurance in Education*, 17(4): 377-392.
- Dillman, D. A. 2007. *Mail and Internet Surveys: The Tailored Design Method*. Hoboken, NJ: John Wiley & Sons.
- DiMaggio, P. J., & Powell, W. W. 1983. The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American sociological review*, 48(2): 147-160.
- Evans, M. 2004. *Killing Thinking: Death of the University*. London: Bloomsbury Publishing.
- Fereday, J., & Muir-Cochrane, E. 2006. Demonstrating rigor using thematic analysis: a hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5: 1-11.
- Galbraith, C., Merrill, G., & Kline, D. 2012. Are Student Evaluations of Teaching Effectiveness Valid for Measuring Student Learning Outcomes in Business Related Classes? A Neural Network and Bayesian Analyses. *Research in Higher Education*, 53(3): 353-374.
- Guest, G., MacQueen, K. M., & Namey, E. M. 2012. *Applied thematic analysis*. Thousand Oaks, California: Sage.

- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. 2006. *Multivariate data analysis*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Hajdin, G., & Pažur, K. 2012. Differentiating between Student Evaluation of Teacher and Teaching Effectiveness. *Journal of Information and Organizational Sciences*, 36(2).
- Hartfiel, G. 1976. Bürokratie, *Wörterbuch der Soziologie*: 103-104. Stuttgart: Alfred Kröner Verlag.
- Hood, C. 1991. A PUBLIC MANAGEMENT FOR ALL SEASONS? *Public Administration*, 69(1): 3-19.
- Kallio, K.-M., & Kallio, T. J. 2012. Management-by-results and performance measurement in universities – implications for work motivation. *Studies in Higher Education*, 39(4): 574-589.
- Kanter, R. 1977. *Men and Women of the Corporation*. New York: Basic.
- Kristensen, J. E., Elstrøm, K., Nielsen, J. V., Pedersen, M., Vind Sørensen, B., & Sørensen, H. 2007. *Ideer om et universitet*. Aarhus: Aarhus Universitetsforlag.
- Kristensen, J. E., Nørreklit, H., & Raffnsøe-Møller, M. 2011. Introduction: University Performance Measurement at Danish Universities. In J. E. Kristensen, H. Nørreklit, & M. Raffnsøe-Møller (Eds.), *University Performance Measurement: 7-17*. Copenhagen: DJØF Publishing.
- Lindahl, M. W., & Unger, M. L. 2010. Cruelty in Student Teaching Evaluations. *College Teaching*, 58(3): 71-76.
- Luft, J., & Shields, M. D. 2003. Mapping management accounting: graphics and guidelines for theory-consistent empirical research. *Accounting, Organizations and Society*, 28: 169-249.
- Lyotard, J.-F. 1984. *The Postmodern Condition: A Report on Knowledge*. Manchester: Manchester University Press.
- Merton, R. K. 1940. Bureaucratic Structure and Personality. *Social Forces*, 18: 560-568.
- Merton, R. K. 1957. *Social Theory and Social Structure*. Glencoe, IL: Free Press.
- Moed, H. F., Burger, W. J. M., Frankfort, J. G., & van Raan, A. F. J. 1985. The use of bibliometric data for the measurement of university research performance. *Research policy*, 14(3): 131 - 149.
- Moullin, M. 2007. Performance measurement definitions. *International Journal of Health Care Quality Assurance*, 20(3): 181-183.

Neely, A., Adams, C., & Kennerley, M. 2002. *The Performance Prism: The Scorecard for Measuring and Managing Business Success: The Scorecard for Measuring and Managing Stakeholder Relationships (Financial Times Series)*. Edinburgh Gate: Pearson Education Limited.

Otley, D. 1999. Performance management: a framework for management control systems research. *Management Accounting Research*, 10(4): 363-382.

Parker, M. 2002. *Against Management: Organisation in the Age of Managerialism*. Oxford: Polity.

Platt, M. 1993. What student evaluations teach. *Perspectives on Political Science*, 22(1): 29.

Raffnsøe-Møller, M. 2011. Aims and Formats for performance measurement at Danish Universities. In J. E. Kristensen, H. Nørreklit, & M. Raffnsøe-Møller (Eds.), *University Performance Measurement*: 49-78. Copenhagen: DJØF Publishing.

Rosanas, J. M., & Velilla, M. 2005. The ethics of management control systems: Developing technical and moral values. *Journal of Business Ethics*, 57(1): 83-96.

Schubert, T. 2009. Empirical observations on new public management to increase efficiency in public research - boon or bane? *ISI - Fraunhofer Institute Systems and Innovation Research. Research Policy*, 38: 1225-1234.

Simpson, P. M., & Siguaw, J. A. 2000. Student Evaluations of Teaching: An Exploratory Study of the Faculty Response. *Journal of Marketing Education*, 22(3): 199-213.

Strang, D., & Macy, M. W. 2001. In Search of Excellence: Fads, Success Stories, and Adaptive Emulation. *American Journal of Sociology*, 107(1): 147-182.

Stöltzing, E., & Schimank, U. (Eds.). 2001. *Die Krise der Universitäten*. Wiesbaden: Westdeutscher Verlag.

Temple, P. 2014. *Universities in the Knowledge Economy. Higher education organisation and global change*. Abingdon: Routledge.

Van de Walle, S., & Hammerschmid, G. 2011. The Impact of the New Public Management: Challenges for Coordination and Cohesion in European Public Sectors. *Halduskultuur - Administrative Culture*, 12(2): 190-209.

Ward, S. C. 2011. *Neoliberalism and the Global Restructuring of Knowledge and Education*. New York: Routledge.

Weber, M. 1905. *Die protestantische Ethik und der 'Geist' des Kapitalismus, II. Die Berufsidee des asketischen Protestantismus*.

Weber, M. 1968. *Economy and Society: an outline of interpretive sociology*. New York: Bedminster Press.

Wren, D., & Bedeian, A. 2009. *The Evolution of Management Thought*. . NJ: Wiley.

Wright, S., & Williams Ørberg, J. 2008. Autonomy and control: Danish university reform in the context of modern governance *Learning and Teaching*, 1(1): 27-57.