

# Who tells fortunes to nonprofit professional theatre producers? State intervention and the strings attached

Ellen Loots

*University of Antwerp, Belgium (ellen.loots@uantwerpen.be)*

Submission for ACEI 2014, Montreal

*In the case of interest in citation, please contact the author for a recent update.*

**Abstract.** The present paper investigates the evaluation process that a patronage-giving state has developed in order to proficiently judge the worthiness of applicants to receive public support. It combines content analysis and regression-based techniques and elicits a discrepancy between the verbal evaluation (the written reports based upon the criteria as defined by the state) and the underlying decision-making process of evaluation panels. *Artistic quality* is the major determinant of subsidy allocations to the detriment of *social concerns*. The absence of more direct effects in our analyses is striking and the conclusion that only a fraction of the imposed criteria matters is conspicuous. We probe some implications for the performance of theatre producers, which need to balance between mission and survival.

## Highlights

**Key words:** Performance, theatre, evaluations, allocation decisions, expert panels

## 1. Introduction

The performance of organizations has been a major yet complex issue in management and organization studies. Back in 1986 Venkatraman and Ramanujam

(1986: 801) already addressed the complexity of defining performance: “*Although the importance of the performance concept (and the broader area, organizational effectiveness) is widely recognized (...), the treatment of performance in research settings is perhaps one of the thorniest issues confronting the academic researcher today. With the volume of literature on this topic continually increasing, there appears to be little hope of reaching any agreement on basic terminology and definitions.*” In the last decades some dominant performance indicators have become salient in the canonical literature on assessing and explaining organizational performance. In the majority of industries goals have a major financial part, because profit maximization is the ultimate objective of most firms. This is realized and corroborated through instances such as optimal production, sales maximization and customer service. But since organizations have multiple and frequently conflicting goals, the definition of performance remains a challenge for researchers in any field of management (Chow, Heaver, & Henriksson, 1994). A clear understanding of the performance of organizations active within the boundaries of a field or industry is not a redundant undertaking, since the improvement of organizational performance is “*at the heart of strategic management*” (Venkatraman & Ramanujam, 1986: 801).

In arts, especially in the arts that are situated at the non-profit end of art production, performance definitions and measures are equivocal. This can be explained partially by the involvement of several stakeholders. Applicable to most art forms is the idea that artistic creativity is not as much an act of the individual genius as it is the outcome of a collaborative effort by many, as expressed by sociologists with Howard Becker as a forerunner in the 1970's (Becker, 1974). Several have studied the critical roles of a range of actors that are complementary to the artist in creating and displaying works of art (Caves, 2000; Hirsch, 2000; Voss, Cable, & Voss, 2000; White & White, 1993 (1965)). It is easy to imagine that the interests and values of these actors not necessarily align. As such, the field of arts or cultural production (Bourdieu, 1983; Dimaggio & Hirsch, 1976; Peterson & Anand, 2004) can be comprehended as one at “*a juncture of different societal logics*” (Goodrick & Reay, 2011: 406). The field of cultural production is one affected by logics of the market (consumer demand) as well as by that of the professions (Greenwood, Diaz, Li, &

Lorente, 2010; Oakes, Townley, & Cooper, 1998), which in arts do not only include artists, but likewise arts managers, distributing companies, critics and other experts. Besides that, also the state and public officials are strongly involved, especially in supporting art that is developed by not for profit organizations and considered having public-good or merit-good characteristics and positive externalities (for a discussion, see Throsby (1994) and Fullerton (1991)). All parties would want an audience to meet the art produced, for sure; but the endogenous dynamics navigating each, such as interests, values, power dependencies, and capacity for action (as categorized by Greenwood and Hinings (1996)) do not necessarily align (Krebs & Pommerehne, 1995). If performance may be defined as the extent to which predetermined goals are achieved (Chow et al., 1994), we can imagine that the involvement of several stakeholders – including taxpayers and their political representatives – may lead to a multifaceted bundle of possibly conflicting performance goals that art producers are confronted with. These goals beyond the control of the producers are outspoken or candid, and they may serve different functions, which extend far beyond the production of high-quality arts. Performance evaluations by external parties may take place for different reasons, including monitoring, evaluating and affecting the behavior of arts organizations (Schuster, 1997).

What are the implications for the producers of arts? Although the primary source of artistic creation is the effort of an individual, the majority of activity in arts production is carried out by firms (Throsby, 1982). On the one hand, these art producing firms can be understood as commodity producers, that generally attempt to convert all actions into the buying and selling of commodities, or converting an audience to their artistic offering (Friedland & Alford, 1991). On the other hand, relying upon revenue grants for their operations, these organizations to some extent resemble public sector organizations that must be organized in accordance with state logics, as one of imposed efficiency (a business-like logic of cost-effective operations) as well as one of meeting particular needs of citizens (Reay & Hinings, 2005, 2009). In democratic states, cultural policy is generally aimed at the provision of art and contended to be contributing to the wellbeing of citizens (Feder & Katz-Gerro, 2012). Policymakers seek to justify the distribution of public means, for

which they develop a rationale and a conditional rule-set. So grants to organizations may have strings attached in that they stipulate how an organization is expected to behave, at least on some dimensions. Thereto, most governments foresee an assessment procedure, that is partially business-based, partially oriented on the artistic output. Public donors have been said to define the conditions for public support in terms of variables that are easily observed, such as capacity utilization (Krebs & Pommerehne, 1995). Nonetheless, these performance evaluations by public donors, both *ex ante* and *ex post* grant allocations, regularly intend to also assess the intrinsic quality of an organization's merchandise (Schuster, 1997) and outcomes rather than outputs (Pignataro, 2011), dimensions which are not only tough to define, but also difficult to measure.

In the present paper, we study the granting decision processes in two democratic regimes in continental Europe: the Netherlands and the Dutch-speaking region of Belgium (Flanders). More specifically, we study the relationship between the conditions to theatre producers as put forward by a state in a set of formal criteria, and the evaluation of these aspects as assessed by expert panels that are prearranged to do so. One of the dimensions under scope is artistic quality, the major concern of many of the grant-applicants because it is their reason of existence. We conceive of these criteria as performance requirements by actors external to the organization itself and external to a market of demand. From this inquiry we infer implications with regard to the performance of producers of art that is not sustained by market demand.

Scholars have regularly examined various aspects of the *coproducing* and *taste making* functions of different types of intermediaries in cultural markets (Glynn & Lounsbury, 2005; Shrum, 1991). In contrast, the decision processes by which these intermediary brokers execute *selections* have gained less momentum, except for some recent contributions (Foster, Borgatti, & Jones, 2011; Seifert & Hadida, 2006; Wijnberg, 2011). The subsidy allocation process can be seen as a decision process which is vital to subsidy applicants, which in some cultural subfields heavily rely on support by the state. It is a process of inclusion and exclusion about which the

government and/or its appointed representatives decide. Due to the dependency of art producers on subsidies, the subsidy allocators gain the power to as gatekeepers safeguard the barriers to entry to, and to cause exit from an art production environment. We extend the literature on gatekeeping in the arts that is often considering behavior by an individual, often in his professional occupation, by advancing a political process as an act of gatekeeping.

We find a clear mismatch between policy and praxis and we probe that the involvement in the subsidy allocation process of two distinct parties, with diverging prerogatives and interests, may introduce an agency problem that underlines the dialectical relation between the art world and politics; in the countries under scope quality-maximizers have bigger odds for receiving subsidies than attendance-maximizers, and art producers have no incentive to embrace a central government's social objectives.

The remainder of the present paper is as follows. After clarifying the method that is used for these analyses, we expose the context of our inquiry. We introduce the policy requirements toward subsidy-seeking arts producers as criteria of an evaluation process. The subsequent steps of our analysis consist of counts, correlation analyses and regression analyses in order to investigate respectively the transparency of the evaluations, the requirements advanced to art producers in order to obtain legitimacy, and the determinants of the reception into state sustenance. In a conclusive section we probe a number of explanations for the processes at hand and we induce implications for arts producers.

## **2. Method**

### 2.1. Abduction

The general method of the present inquiry is in line with what has been referred to as an *abductive logic*.<sup>1</sup> Charles Sanders Peirce suggested the logical category of

---

<sup>1</sup> We deliberately use the expression "is in line with...", because we study just two cases, in which we observe a pattern that repeats itself. According to Schurz (2008: 219) a legitimate scientific abduction would be one in which "the postulated entity or property explains *many intercorrelated* but *analytically independent*

abduction in research. It provides criteria for scientific discoveries distinct from *criteria of proof*. Where deduction is a process of checking theories, and induction one of checking hypotheses, abduction can be conceived of as a method for the discovery of hypotheses ('on probation') or propositions, the latter being abductive explanations of surprising facts. Starting with the observation of a new or surprising phenomenon, abductions serve the goal of inferring something about the unobserved causes or explanatory reasons of the observed events (Schurz, 2008). It is then necessary to add interpretations to the facts, or assumptions about possible interpretations (propositions), in order to adapt them to situations that are different from those in which they have been observed. A valid abductive judgment establishes the possibility of an explanatory concept, rather than supporting a unique law or theory. An abductive line of reasoning may particularly be convenient for social studies that have a main purpose of evaluation and in which attempts are made to give effective explanations for field-dependent phenomena (Levin-Rozalis, 2000).

## 2.2. Quantitative content analysis

We use the written judgments of the evaluation committees as a major source of information, from which theorizing can emerge. Thereto we apply content analysis (Bryman, 2008; Kassirjian, 1977; Kolbe & Burnett, 1991; Krippendorff, 1980; Krippendorff & Bock, 2009; Riffe, Lacy, & Fico, 2005). Content analysis has been defined as "*a research technique for making replicable and valid inferences from data to their context*" (Krippendorff, 1980: 21). More specific, "*Quantitative content analysis is the systematic and replicable examination of symbols of communication, which have been assigned numeric values according to valid measurement rules and the analysis of relationships involving those values using statistical methods, to describe the communication, draw inferences about its meaning or infer from the communication to its context, both to production and consumption*" (Riffe et al., 2005: 25). A close-up upon textual information is salient for this research for at least two reasons. First, we expect that in our context a specific vocabulary is used to

---

phenomena, and in this sense yields a *causal or explanatory unification*." The context-dependence of the phenomena we study is so that the explanations we will be able to advance are neither factual nor exclusive.

evaluate art producers and their products. As in science, we expect instances such as ‘original’, ‘innovative’, ‘important’ or ‘significant’ to express valuation (Guetzkow, Lamont, & Mallard, 2004). Bielby, Moloney, and Ngo (2005) have pointed to the very specific language that is being used to describe quality dimensions of movies, such as ‘subtlety’, ‘realism’, ‘credibility’ and ‘plausibility’, ‘complexity’ (or ‘ambiguity’). A similar vocabulary can be expected for the description of theatrical work. Second, the specific type of this text appeals for a profound approach, given its hermeneutical status. It is a written judgment that expresses the reasons for why an organization should be or should not be supported by the state, guided by a bundle of predetermined criteria. It is a judgment of field members on other field members, who need be able to look into each other’s eyes afterward. So an in-depth reading of these texts is necessary because we expect these judgment reports to evaluate many aspects on different levels of interpretation. These reports are a summary of a discussion and since they enter in the semi-public domain where they motivate a decision which is vital to organizations, we expect them to be accurate in reflecting the argumentation that leads to this decision outcome.

Content analysis implies a quantification of text excerpts. Thereto a coding scheme is to be developed, for which a first step is the defining of content categories. In our study these categories are the criteria set by the government. Since they are short in commentary and have no descriptive clarification<sup>2</sup>, we had to produce a set of reliable and valid indicators that expressed the distinguished content and dimensions of each criterion. Also to determine is the basis unit of text to be selected and classified, which could be paragraphs, thematic units, sentences or words (Strijbos, Martens, Prins, & Jochems, 2006). We decided to use thematic units on the level of the sentence or block of sentences. The selecting of the content and attributing it to a category was an iterative, recursive and reflexive process. Independently two researchers coded the texts making use of Nvivo qualitative research software. During the analysis ‘dictionaries’ were developed. These consisted of lists of words and phrases that were associated with the categories of the content we looked for.

---

<sup>2</sup> Except for the artistic quality variable in the Netherlands

After the coding the results were pooled and the final retentions were discussed and agreed upon. These dictionaries can be retrieved upon request.

The coded texts led to numeric variables. For each organization two dummy variables were developed. A first variable v1 indicates the presence (1) or absence (0) of a topic on the level of the organization. It expresses whether the evaluation panel assessed an organization to a specific criterion, or if it neglected to do so. A second variable v2 is a valued judgment: we assigned the value of one when the panel was positive about an organization on a singular dimension and the value of zero when the panel was negative or not pronounced.<sup>3</sup> Covering multiple dimensions, the descriptions of a company's artistic quality were thick and nuanced, which led us to code 'artistic quality' into a more fine-grained scale of 5 values. The Average Deviation Index by Burke & Dunlap (2002) measures the dispersion of responses around the mean; we obtained a value of  $AD_{M(J)} = 0.43$ , which can be considered indicative of acceptable inter-rater agreement.<sup>4</sup> If not similar, the scores of the two raters were averaged.

### 2.3. Data

We study the evaluations of 84 Dutch and 57 Flemish theatre organizations that applied for a subsidy for the four-year term following the application years 2008 and 2009 respectively.<sup>5</sup> Choosing these two areas had several advantages. First, the political system, policy priorities and legislation bear similarities, which makes comparisons salient. Second, the Netherlands and Flanders in Belgium share the Dutch language. From a methodological point of view this implies that content

---

<sup>3</sup> The value of 1 can thus be interpreted as the panel considering a quality or activity of an organization as a defensible argument in favor of state support, so as a flag going up allowing an organization to continue based on the criterion under consideration. In contrast, a negative evaluation on a criterion or the absence of a positive evaluation both imply that a panel is not convinced of the qualities of an organization on the term of focus.

<sup>4</sup> An  $AD_{UL}$  value (UL = upper limit) for 5-point items scales is 0.8 (the cutoff value of 0.8 is statistically significant at the .05 level for all Likert-type scales with five or more response options, as long as the sample size N equals or exceeds 13) (Burke & Dunlap, 2002).

<sup>5</sup> Belgium is subdivided into three language-based communities that organize some dimensions of policy separately, such as culture and arts. We focus on Flanders, which is the Dutch-speaking northern region, neighboring to the Netherlands, with about six million inhabitants. We study the previous term, based on decisions of 2008-2009 (because the most recent evaluations (2012-2013) were not available at the time of the analyses).

analyses could be performed and interpreted without the intermediate step of translation from one language into another, or to English (with the risk of during the process losing the subtleties of the language). From a market point of view this implies that mutual exchange of theatre supply is extant, leading to commonalities in theatre traditions. Third, the population of theatre organizations applying for subsidies is in both cases sufficiently large to allow for statistical analyses, yet not too outsized for content analysis. A similar study of the dance companies or operas within these countries would suffer from a lack of observations, whereas larger pools of applicants as in larger countries may prevent the in-depth approach required for text analyses.<sup>6</sup>

The theatre companies under study can be divided into two distinct groups, of touring companies and companies that reside in their own playhouses.<sup>7</sup> As in other continental European countries, structural state support is intended to contribute to the operations by filling-up the yearly deficit, which can mount up to a share of 70 to 80% of turnover. Kindled by the New Public Management movement the public sectors in the Netherlands and Belgium have introduced performance management systems and open communication on policy decisions, budget allocations and outcomes. In both countries the application and assessment process under scope takes place every four years. Theatre companies and theatre houses can submit an application file that is being assessed on its financial and administrative plans, and on its artistic plan. For the latter expert panels are invoked. These panels consist of individuals that are attributed an institutional responsibility to make judgments on the artistic performance in arts. Panel members originate from within the country and are seen as experts in the domain. The Dutch panel consists of eight members, who all display an active involvement in one or more theatre houses or groups, either artistically or managerially: one actor, a dramaturge, a director, one theatre programmer, and three managers and one head of finance. With twelve members the

---

<sup>6</sup> As an example for a small number, in the Netherlands 23 organizations received subsidies as a dance company for the term 2005-2008, compared to 86 theatre organizations; In Italy there were 819 applicants for funding of activities in music in Italy in 2005 (as mentioned by Zan, Bonini Baraldia, Ferria, Lusiana, & Mariana, 2012).

<sup>7</sup> The latter also travel with their productions. After World War II the central government in the Netherlands started to financially contribute to theatre production, which used to be the exclusivity of municipal entities, but only conditional upon traveling activities. The underlying idea was that the economic and moral resurrection should not remain a privilege to metropolitan areas (OCW/Boekmanstudies, 2007).

panel in Flanders is larger. Four theatre critics, three programmers, two theatre scientists, one scenarist, a television producer of arts programs and a museum director together compose a panel, the members of which are not employed by a theatre house or company.

Members know the criteria to use, but they are generally not instructed on how to interpret the specific criteria or on how to ponder each one. They are only accountable to other panel members and no rules on how to evaluate are written down. Granting decisions fall in a meeting: no individual rating preceding this gathering, throughout a discussion panel members arrive at a consensus which is being reported in the documents we study. Members leave the room whenever organizations they are involved with are topic of discussion.

The policy program goals are set in legislation and manuals, which also inform the applicant pool. The criteria against which they are assessed ground in these policy documents. In Flanders the Arts Decree stipulates the political priorities related to performing and visual art forms, and the conditions under which organizations can apply for government support in the form of direct grants. A handful criteria is related to structural aspects of the applicant, such as legal personality, operating base and minimum staffing (Arts Decree, subdivision II, article 7, §1, 1 to 11). Additionally some minimum standards are being postulated, such as a minimum number of activities, a minimum share of market income, and a minimum share of turnover to be spent on wages (Arts Decree, subdivision II, article 7, §2). If compliant to these formal requests, an applicant needs to give proof of acting in accordance with a bundle of major substantive criteria (Table 1). In the Netherlands a corpse of funding bodies or art councils execute a major part of national cultural policy by dispensing subsidies to organizations and individuals. The Dutch Funds for the Performing Arts has the task to realize a diverse, qualitatively high and well-dispersed supply, which it in the time window under scope developed under the motto ‘more for less’, implying a stringent selection on the one hand, and an increase of the average subsidy on the other hand. The Dutch regulation stipulates that applications are being judged upon five main aspects, of which some of the content is elaborated. In general, the judgment on the quality is based upon *‘the quality of the*

*previous work and/or activities, the qualitative development of the work and/or activities, and the plan and budget.*' The first quality-criterion refers to the artistic quality. Three central notions to the evaluation of this artistic quality have been put forward and made explicit: craftsmanship, originality and expressive power. Opposed to artistic quality, criteria 2 to 6 are being labeled the 'functional quality'. This relates to the way in which productions and performances function within the Dutch society.

Table 1 depicts the resumed dictionaries of the criteria as they have been used to do the content analysis of the judgments.<sup>8</sup> The Flemish criteria were hardly elaborated; the interpretation came into being during the analysis. The Dutch content is largely based upon the descriptions provided within regulatory documents.

### *Table 1*

Three major performance categories underlie the criteria and can thus be perceived as how a state conceives of the performance of arts producers: artistic quality, a social dimension, and a dimension which clusters entrepreneurship and strategic and operational requirements. David Throsby (1982:246) has postulated that an arts funding decision can be comprehended as one "*involving essentially two maximands: participation rates, and the quality of the output.*" Attendance here is not just understood as a magnitude, but also invokes a certain level of diversity of the audience, including socially weaker groups. The artistic and social dimensions of the art production are clearly paramount; the inclusion of criteria that specify requirements of efficiency and effectiveness may signal the vigilance by the government on the appropriation of state money. In an extensive literature review of performance in the creative industries Hadida (2013 (forthcoming)) recognizes three core dimensions of performance: commercial performance, artistic merit and societal

---

<sup>8</sup> These are summarizing versions of the full dictionaries, which can be retrieved upon request. Especially the description of the artistic quality variable has been curtailed.

impact. A fourth dimension is present in the literature, yet to a lesser extent: Hadida labels it ‘managerial performance’, referring to “*creative managers’ commitment to and effectiveness in the execution of their functions*” (2013 (forthcoming): page). These performance dimensions can be clearly recognized within the requirements that are advanced by an evaluator external to the organization. Commercial performance here is not at stake, because the subsidy system for theatre production requires the non-profit character of applicants. The achievement of artistic quality and social objectives clearly underlies the Dutch and Flemish governments’ cultural policies. Both targets are projected onto grant-receiving organizations that are committed to realize as such, giving proof of an efficient and strategic management.

### **3. Results**

#### 3.1. The comprehensiveness of the evaluations: counts

Theoretically, panels can be expected to assess organizations on each single criterion imposed by the government, but in practice evaluations might not happen as formal and tight. This section studies the occurrence of criteria that are being put forward by the government (variable v1). Table 2 displays the incidence of the various criteria within the judgments.

#### *Table 2*

##### 3.1.1. The occurrence of the criteria

Each organization is in each of both countries under scope being evaluated on its *artistic quality*, which may not come as a surprise. Although both governments put forward a dozen criteria to which applications have to conform, panels select subsets of criteria to evaluate single organizations. *Reputation*, which is connected to past performance, is the second most heavyweight criterion in terms of occurrence in Flanders (91%). This could imply that newcomers might have difficulties entering the subsidy system. Whether or not a plan is *feasible* is being reported in 85% of the

Flemish cases. This may be a decisive criterion. Yet, we could question the salience of this criterion, because the panel only evaluates an artistic plan, and not its financial counterpart. Other criteria that are frequently invoked in Flanders are *profile and position* (81%), and *audience orientation* (84%). *Position* implies many distinct dimensions such as a geographical, unique, prominent, pioneering, or visible position. Also in the Netherlands some distinct dimensions of *profile and position* (100%) come to the fore. A subset of organizations is being assessed on its uniqueness or distinctiveness. In the Netherlands the expression ‘complementary’ is being used; in Flanders the notions ‘added value’ and ‘significance’. In addition to that, there appear to be some strong actors (important, remarkable key players), and some exemplary (or innovative or pioneering) actors. Two criteria remain underrepresented in the Flemish evaluations: *long term vision* (38%) and *realism of a growth path* (19%). It appears as if features that are related to the content of the grant application, often projecting a future plan, do not prevail to an organization’s reputation, which is mainly built upon past performance. This finding is in line with Gorman (2007) and Guetzkow et al. (2004) that illuminated similar tendencies within evaluations in science, referred to as the ‘Oppenheim effect’.<sup>9</sup> At the tail of the list of criteria in Flanders (38%) and not an official Dutch criterion (but yet mentioned), is *vision* (31%). A close reading of the texts suggests that in both countries a lack of vision may be a sufficient condition for rejection.

In sum, we see that in Belgium there is more consistency in the application of criteria, even though they are manifold, and open-ended. In the Netherlands where fewer criteria are to assess, less perseverance in their application can be noticed. Profile and position are austere being considered in these assessments, giving the impression that panels perceive as its major task to compose an artistically diverse landscape of complementary organizations. Yet, the evaluation procedure across organizations is very unsystematic, almost ad hoc.

### 3.1.2. Trust as an alternative evaluation practice?

---

<sup>9</sup> The Oppenheim effect in the peer reviewing process by scientific journals describes the phenomenon that the author and not the manuscript itself is the criterion in determining quality to the submission and is the base for the editorial decision (Gorman, 2007)

Taking into consideration two distinct entities, namely a country and a region, advances some insights in the peculiarities of each of the evaluation processes. We mark that the Dutch panels explicitly judge the trustworthiness of more than half of the applicants (n=43), whereas the Flemish evaluations are far more factual. As an example, the Dutch panel may have trust in a theatre company proceeding in its development, or in a new artistic leader as a drama producer. Or it may distrust an organization, as: *“In its future plans the company does not advance a clear vision on the artistic development or the themes it would like to treat. That is why the committee does not have sufficient trust in the expected quality of the plays that the company would want to produce in the upcoming years.”*

Kee and Knox's (1970) interpretation of trust as phrased by Mayer, Davis, and Schoorman (1995) refers to *“the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party.”* Trust is not about taking a risk per se, but rather it is about a willingness to take risk (Mayer et al., 1995). For someone to trust a person to produce some beneficial event, he must have confidence that the other individual has the ability and intention to result (Deutsch, 1960), or, as refined by Mayer et al. (1995), he must hold the perception that the trustee is trustworthy, which he could be by giving proof of ability, integrity and benevolence. Ability is the bundle of skills, competences and characteristics that make someone able to achieve a specific goal or to have influence within a specific domain (Mayer et al., 1995; Zand, 1972). Obviously, in the process at stake of the assessments of artistic organizations, the ability of producing artistically significant work is a central criterion. It is not always easy to argue rationally why a theatre producer is of high quality, or believed to perform well when talking about artistic realizations. Trust as an emotional expression of worthiness - as an evaluation criterion is applied in over 50% over the evaluations, and strongly correlates to the subsidy allocation decision (corr. = 0.80). So in the Netherlands, the judgment is likely to depend on the trust the panel has in the artistic quality as it is expressed in the plan, or as it was displayed in past performance. Trustworthiness can not only be perceived of as an additional criterion

or an attribute of an actor, but as a specific *modus operandi* in the Dutch evaluations. This emotion-driven way of evaluating appears in sheer contrast with the process in Flanders, where panels attempt to find the adequate vocabulary to rationally defend their decisions. It is not clear whether this phenomenon is random, or whether cultural differences or differences related to the composition of the panel<sup>10</sup> may explain these distinct approaches.

### 3.2. Features of legitimate actors: correlations

The correlations between the variables (v2) will clarify two things. Correlations between criteria and judgments will first reveal which criteria appear to be counting in the decision making process. Second, significant correlations among criteria will disclose which criteria relate and might co-occur within the assessments. If legitimacy can be understood as the outcome of being evaluated as doing the appropriate things, these correlation coefficients give insight in the characteristics and activities of legitimate actors in the theatre fields in Flanders and the Netherlands. Many scholars have written about legitimacy and its different forms, antecedents and consequences. We here conceive of legitimacy as it was defined by Suchman (1995), as moral legitimacy, which is a form of legitimacy based on the normative approval by an expert audience (Suchman, 1995).

#### *Table 3 + table 4*

##### 3.2.1. Correlations between the criteria and the outcome variable

The correlation between *artistic quality* and the judgment is highly significant. This is the case in both countries, with the Netherlands (0.887) outperforming Flanders (0.778) on this issue. This finding implies that the assessments are a clear reflection of the artistic quality of an organization, as it is perceived by a group of experts

---

<sup>10</sup> Recall that in the Netherlands the panel consisted out of people that themselves work for art producing organizations, whereas the Flemish panel was formed by people external to art producers, such as critics.

entitled by the state to assess drama organizations on their performance. Yet the correlation is not perfect; some unexplained variance of 10% to 20% remains. This suggests that some judgments are advanced despite of an artistic evaluation in one or the other direction and that other (latent) factors may be at stake. A remarkable difference between the two countries relates to *audience orientation*. It is significantly though moderately correlated to the judgment in the Netherlands (0.463), while in Flanders any relationship lacks (0.115 not significant (n.s.)). This lack of relationship could indicate that a large or specific audience base does not have any account in the subsidization of theatre in Flanders. Another criterion that both states apply is *diversity*. The diversity focus in Flanders is somewhat ambiguous in its relation to the outcome, whereas in the Netherlands, albeit less taken into account, it is significantly correlated to the arbiter's judgment (0.232 n.s., 0.385, resp.). The criterion of *feasibility* is correlated to the judgment, even on a fairly high level (0.576). Two other criteria that might reflect upon future undertakings of Flemish organizations, *vision* and *realism*, are more contested in relation to the evaluation outcomes (0.295 & 0.331, moderate significance), as is *collaboration* (-0.281). Collaboration shows a negative sign, suggesting that collaborations would be counterproductive in order to receive a positive assessment. This, however, is not in line with political inclinations by the Flemish authorities that instigate on collaborations in order to keep down the expenditure to production. Likewise in the Netherlands a claim for effectiveness has been converted into evaluation standards: for an organization to be subsidized, it should express *cultural entrepreneurship* (0.485) and inscribe within the Dutch "*chain idea*" (0.253), a principle that tries to organize the field in an efficient and cost-effective way. The latter is only moderately correlated with the judgment, as opposed to *cultural entrepreneurship*. At last, two substantive priorities of the concurrent policy maker in Flanders give the impression of not being considered by evaluators. An actor's *social* and *regional value* can be considered to be political values that do not align with an expert committee's priorities, be it by a lack of interest or a lack of knowledge or another latent motivation (0.106 & 0.046, n.s.).

### 3.2.2. Pairwise correlations between the criteria

Looking at the pairwise correlations between the criteria, we see that in both countries three criteria significantly correlate: *artistic quality*, *position* and *reputation* (0.572; 0.616 & 0.566). This is not an unexpected finding, since artistic quality could be an imperative determinant of one's reputation and the position one uptakes. Since we are not considering pure facts, but assessments by one party, the relation may be not so sensible. Hence, we are looking at perceptions of quality, reputation and position, and these perceptions might be entangled. If someone is convinced of a party's quality, he or she could assume that its reputation must be good, or that it holds a relevant position. Or being convinced by the qualities of an organization, someone could even claim that this organization has a good reputation without taking into account external elements that support this prerogative. In Flanders, an additional criterion adds to this triangle: feasibility (values between 0.455 & 0.533). The relation is less straightforward. An evaluating body could obviously trust upon previous achievements by an organization, to gauge whether it will be able to deliver what it projects. But, as advocated earlier, feasibility could be a key determinant, functioning as a barrier to entry or exit from the subsidy system. In addition to scoping upon the presence of correlations, it is also salient to look at the absence of correlations. In Flanders, the criterion of *regional value* does not relate to any other variable. This might give proof of a discrepancy between a (mere) regional value, and the worthiness of grant allocation: grants may be privileged to organizations that attain such a level of overall quality that they perform beyond regional or local latitude. Also criteria that relate to *audience orientation*, *social orientation*, and *diversity*, all recent political priorities do not coincide with values that are typically associated with the high arts: *quality*, international appeal or *reputation*, and *position*. These findings suggest that the value systems or norms of expert panels and grant allocators do not align. The nature of the discrepancy recalls the type of division that Pierre Bourdieu pointed to several decades ago (Bourdieu, 1979).

### 3.3. The determinants of state sustenance: regression-based analyses

A final statistical estimation will elicit whether evaluation panels actually invoke the criteria that are advanced by a state. In theory this should be the case, because the

funding bodies present them to applicant-organizations as a non-exclusive list.<sup>11</sup> As a consequence, we would expect significant coefficients of a similar magnitude for all the variables in the model. The correlations however suggest that only a subset of criteria affects the panels' decisions. Still, the written evaluations give the impression that the panels assess organizations on several dimensions, certainly in Flanders. Yet, there could be a discrepancy between the rhetorical argumentation of the panel, and its decisive behavior. In this last section, we regress the different criteria against the outcome variable, which is the decision of the panel.<sup>1213</sup>

### 3.3.1. Econometric analysis of the Dutch dataset: Firth logistic regression

In the Netherlands, the evaluation committee expresses its evaluation straightforward, as a 'yes' or a 'no'. We can conceive of this polar two-category outcome as a dichotomous dependent variable  $Y$ , with a score  $Y = 1$  if the organization is being evaluated as positive, thus successful, and with a score  $Y = 0$  if the organization receives a negative report. What we actually observe is group membership of each organization (case (1)/ noncase (0)), but what we predict in our model is the population probability  $\pi_i$  of being a case (Cohen, Cohen, West, & Aiken, 2003). Logistic regression is specifically designed to predict the probability of an event occurring (Hair, Black, Babin, Anderson, & Tatham, 2006), which in our case is an appropriate end. For the analysis of the Dutch data we develop a *multiple logistic regression* model that describes the probability of being evaluated positively by a committee of experts as a function of the criteria advanced by policy makers. A logistic regression model is predicting the logit of  $Y$  from  $X$ , that is, the natural logarithm ( $\ln$ ) of odds of receiving either or not a positive evaluation and writes as follows:

---

<sup>11</sup> Although in the Netherlands 'artistic quality' is advanced as the major criterion.

<sup>12</sup> Our samples consist of 84 observations in the Netherlands and 57 in Flanders, which in both cases cover the full population of subsidy applicants. The criteria used for assessment are respectively seven and eleven. This implies that we are dealing with 12 and 5 events per predictor variable (EVP) respectively. These ratios are rather low, but as shown by Vittinghoff and McCulloch (2007), discounting of results from any model with 5–9 EPV does not appear to be justified. This rules out the clustering of predictor variables (as based on a Factor Analysis), which would not square with our intents.

<sup>13</sup> We deliberately do not include external controls. It is not our intention to explain or predict who receives subsidies, and based on which characteristics; our aim is to provide with an insight in whether panels apply the prescribed criteria. Our regression models treat applicants as the units of observation; but our units of analysis are the criteria.

$$\text{logit}(Y) = \ln \frac{\hat{p}}{(1-\hat{p})} = (B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k)$$

Normal logistic regression analysis is not suitable for our dataset.

### **Figure 1 + figure 2**

The plot of the Dutch data suggests quasi-complete separation (figure 1). Hence, only at a ‘low’ value of artistic quality ( $X = 2$ ) there is some overlap between positive and negative judgments, implying that when artistic quality is attributed the value ‘low’ ( $X=2$ ), it does not perfectly predict the outcome of the assessment, contrary to all other quality evaluations (a value of 1 would lead to a negative assessment; values of 2.5 to 5 to positive evaluations). For reasons of comparison we display a similar plot for the Belgian data. Here we dispose of an ordinal ranking of the evaluations, implying more categories. Figure 2 shows that here the phenomenon is absent.<sup>14</sup> Likewise the other variables of the Dutch dataset (not visualized) do not appear to suffer from separation.

In a recently published comparative simulation study based on real-life data sets it was shown that Firth’s penalized maximum likelihood estimation outperforms alternative methods as maximum likelihood and exact conditional logistic regression (Heinze, 2006). Running this corrective method in R by the package *logistf* (Ploner, Dunkler, Southworth, et al., 2010), we find that from all the variables only the effect of *artistic quality* is significant up to the 0.001-level; the effect of *reputation* is just below the threshold of 0.05, also in a second more parsimonious model; no other variable is significant (table 5).

### **Table 5**

---

<sup>14</sup> The plot shows the artistic quality variable. Also the variables position and reputation (not visualized) that due to their relatively high correlation to the judgment could be suspicious, do not appear to lead to quasi-nonconvergence.

### 3.3.2. Econometric analysis of the Flemish dataset: Ordinal logistic regression

Recall that the Belgian assessors expressed their evaluation on a scale, as opposed to the Dutch evaluation, which simply expresses a go-ahead or a halt. Ordinal logistic regression, an extension of the simple logistic regression model, is appropriate for analyses of this kind. The estimation of an ordinal logistic model entails the estimation of a model for the probability of membership in a particular category, based upon values for the independent variables  $X$  and values of the discrete thresholds  $Y_j$ . It models the cumulative probabilities  $P(Y_i \leq j)$  rather than the specific category probabilities  $P(Y_i = j)$  as in binary logistic regressions (Kutner, Nachtsheim, Neter, & Li, 2005). It writes as follows:

For  $j = 1$ ,

$$\hat{p}_{(Y_i \leq 1)} = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$

And similarly for  $j = 2$ ,

$$\hat{p}_{(Y_i \leq 2)} = \frac{e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k)}}$$

and so on for each value of  $Y$ .

A problem related to data separation, with similar symptoms, is *zero cell count*, which occurs when the dependent variable is invariant for a value of a categorical independent variable (Menard, 2002). Figure 2 shows that this is actually the case for the value 0.40 for the Artistic Quality variable (corresponding to value 2 on a 5-point-scale). For ordinal variables (as our AQ-variable) this is not so much of a problem, because a pattern underlying the values is being assumed and used to fill in the blanks in the contingency table that relates the dependent variable to the independent variable. For categorical variables such a pattern cannot be assumed. Plotting the independent dummy variables against the dependent variable, we notice some non-existing X-Y combinations (e.g. position (1) versus a judgment of 1 or 2; no social concern (0) vs. judgment 3..., not illustrated). If the major concern of the

analysis is the overall relationship between a set of predictors and a dependent variable rather than individual values, as is the case, Menard (2002) suggests to proceed and to accept some uncertainty about the coefficient values. Running the PLUM-command in SPSS, we obtain the following function (for threshold  $j = 1$ ):

$$\text{Predicted logit of } (DV \leq 1) = -3.963 + 11.669*AQ + (-2.373)*POS + 1.166*REP + (-0.485)*REGION + (-0.327)*AUD + (-0.367)*DIV + (-0.558)*SOCIAL + (-1.796)*REAL + (-0.918)*FEAS + (-0.797)*VIS + 0.313*COLL^{15}$$

Three variables appear to be significant for predicting the outcome: *artistic quality*, *position* and *realism*. A second parsimonious model only including these three variables exposes that the significance of ‘realism’ dissolves. Compared to the full model of all 11 criteria model 2 explains relatively little additional value, adding only around 3% to the Nagelkerke pseudo  $R^2$ .

### *Table 6*

## **4. Discussion and conclusion**

The evaluation and measurement of the performance of arts organizations, which are subject to accountability toward several stakeholders, is a task fraught with difficulties (Hadida, 2013 (forthcoming); Turbide & Laurin, 2009). Evaluation criteria of different parties do not necessary align and may be tacit. The present paper investigates the evaluation process that a patronage-giving state has developed in order to proficiently judge the worthiness of applicants to receive public support. Our objective has been twofold. First, we explored artistic performance, while conceiving of it as a multifaceted concept that in a field as that of the arts may be influenced, maybe even piloted, by several actors external to arts producing organizations

---

<sup>15</sup> The negative sign for most of the coefficients is explained by the reference category of Y, which is the highest value (category 5). The threshold coefficients are usually not interpreted individually. They specify the points in logit odds where the observations would be predicted into an adjacent category.

(Becker, 1984). Part of our focus was on the relation between *artistic performance* and *artistic quality*. Second, and more specific, we sought to clarify the relationship between performance requirements put forward by a state and the assessments by expert panels that are prearranged to evaluate organizations against formal criteria. This procedure can be considered as an instance where two distinct logics congregate: a better understanding of the underlying mechanisms could give insights in *how* a state logic and a professional logic are able to co-exist.

The application of a mixed-method approach in an abductive line of reasoning appeared prolific to our objectives. By combining content analysis and statistical analysis we elicit a discrepancy between the verbal evaluation (the written reports) and the underlying decision-making process of panelists. Our analyses show that a report containing an assessment of an actor against distinct criteria does not imply that each of these criteria matters in the final decision of the assessors. The redundancy of criteria that reflect social objectives has been illuminated in both Flanders and the Netherlands. A state may expect arts organizations to break down barriers in society and to concern with access maximization and social inclusion. However, for expert panels that by their advices decide about grant allocation, ‘artistic quality’ is a major determinant in the performance assessments, to the detriment of social concerns. Werck and Heyndels (2007) have pointed to an evolution in Flanders taking place between the years 1980 and 2000, of theatre gaining a strong artistic reputation, while attendance has declined. In the Netherlands the number of theatre performances increased with 100% between 1965 and 2002, while participation dropped with 40% (van Klink, 2005). We here reveal that, only some years later, although policy may have tried to impose on theatre organizations to concern about attendance and the inclusion of diverse groups, the procedure is not effective. Expert panels in both countries come to the fore as clearly favoring quality-maximizers over attendance-maximizers. In the Netherlands the ‘artistic quality’ variable is even a predictor of such magnitude that it produces a problem of *quasi complete data-separation*, which was overcome by a *Firth logistic regression model* (Heinze, 2002).

As said, in the cases under scope *artistic quality* is the major determinant of subsidy allocations. Additionally *reputation* and *position* play a role. As conditions that determine the worthiness of applicants to receive state support, both criteria have limitations. First, they are not conceptually distinct from quality. Second, *reputation* is based solely on the perception of panelists. In order to avoid perception biases due to limited cognition, or even favoritism, more objective reputation indicators could be used. A third limitation is that *position* gave proof of being a multi-faceted concept. Panelists evaluate position haphazardly with regard to products and geographical markets; yet, applicants may have a unique position, due to either their distinctive products, either their unique presence in an otherwise deficient region. We advance that peer panels may be capable of determining the first dimension (originality), but that they lack the know-how of judging the second (dispersal). Several elements regarding the nature of the demand side of the market (as household preferences, customary travel distances and the supply in neighboring areas) should be acquainted with in order to gauge the complementarity of the cultural supply.<sup>16</sup> So the proficiency of ‘expert panels’ to judge about the complementarity or redundancy of distinct organizations in this or any other context is disputable.

Our investigation further accentuates a paradox in the evaluation process, which has been illuminated elsewhere: a plan is supposed to be decisive for grant allocation, but past performance reflected in instances such as artistic quality, position and reputation overrule future intents (Gorman, 2007). An undesirable consequence could be that incumbents are privileged, and eventually that innovations are hampered, which is not a-typical for these processes (Zan et al., 2012). We probe that past performance is an unavoidable yet strong determinant of decisions and selections in evaluation processes that primordially rely on the expertise of peer-panels.

---

<sup>16</sup> The country and region under scope are relatively small in surface; and our dataset of drama producing organizations combines distinct profiles, with larger municipal theatres on one side of the spectrum, and smaller touring companies on the other side.

The importance of the present findings lies primarily in the framing of an evaluation process that is developed by one party – the state –, yet executed by a third party – an expert panel –. The absence in our analyses of more direct effects is striking and the conclusion that only a fraction of the imposed criteria matters is conspicuous. Policy makers do not assume panels to assess different organizations with distinct measures and weights. Despite the existence of overt evaluation criteria, the performance evaluation in se is not transparently based on them. Of course, the present study is limited because it empirically investigates a particular evaluation process of one artistic sector at a given date in only two countries. Still, we can advance some tentative explanations (hypotheses on probation) and implications for art managers that may be rooted in two distinct agency problems.

A first meaningful relationship is that between the government and art producers. The production and distribution of theatre plays has become one that is organized by private parties, most of which non-profit. Central governments as in Netherlands and Flanders engage in subsidiary relations with these producers, as to provide civilians with a cultural supply. Implicitly, this strategy reflects the premise of arts as a public good, or as one that generates externalities, which justifies why a government may sustain theatre supply. Yet the relationship between the government (principle) and the art producers (agents) is not without glitches. When the desires or goals of the principal and agent conflict and when it is difficult for the principal to verify what the agent is actually doing, an agency problem arises (Eisenhardt, 1989). The two conditions hold here. The splits between a cultural policy that fosters the quality of the artistic production and that simultaneously engages with participation has been a constant, and although the object of recurrent debate, policy makers rarely (if ever) succeed in bridging the gap between these two goals, let alone in developing adequate instruments that serve the linkage of these objectives (van Klink, 2005). This may be a germane political problem that is passed on to producers of art that allegedly prioritize the quality of their creations. The goals of the principle and agent may not just be conflicting; policy goals may even so be unclear, due to phenomena such as recognized by van Klink (2005) in the Netherlands as *policy inflation* (the constant renewal or rephrasing of policy aims, due to the fact that every four year

new politicians seem reluctant to endure the work of their predecessors) and *process inflation* (the increasing bureaucratization and complexity of policy processes for the sake of transparency). Further, the monitoring of arts providing agents appears to be a difficult process, leaving the principle in uncertainty about the agents' actions and their contributions to his objectives. This may be indicated by the mismatch between the goals of the principle (quality and attendance maximization), the instrument (four-year subsidy after a plan-based evaluation) and the processor (a panel of artistic experts and peers). Artistic quality is hard to define, and the information that is used for allocating public means is a written-word commitment of an organization which may be a meager basis to build a contractual relation on. This document serves as the basis for a performance evaluation by expert panels supposedly proficient in doing so. As we have shown, these panels mainly rely upon realizations that art producers have proven to be capable of in the past, they tend to neglect social and attendance concerns, and they assign public means based on their trust or mistrust in an applicant. If goals of subsidy providers would be more crystallized, measurable performance outcomes could be advanced. The recognition of the path-dependency of both artistic performance and cultural policy could open the door for an alternative support mechanism less subject to subjectivity and arbitrariness.

A second agency relationship can be recognized in precisely this relationship, between policy makers and the expert panels that execute the evaluations. Also here an agency problem abounds, due to goal conflicts and the lack of transparency of the evaluation processes. These two actors -the state and the professions- have been identified as the primary shapers of institutional forms (Dimaggio & Powell, 1983). The concept of institutional logics allows to address more clearly issues concerning different models of organizational functioning and accountability (Friedland & Alford, 1991). The logics that powerful actors such as the state (the principle) and the professions (the agent) attain may be distinct, but also able to co-exist (Goodrick & Reay, 2011). Our case of the assessment of the performance of organizations in arts is exemplary for systems in which two actors, that here in a dyadic relationship execute resource allocation, do not espouse similar institutional norms nor share the same interests, similarly to what has been described elsewhere (Scott, 1987;

Townley, 1997). Although public authorities set the directions for granting decisions, our analyses clearly show that the mere evaluations deviate from formal and overt conditions, which suggests a goal conflict. As the quality of the artistic production is the major determinant for public support, which can be subsumed to be the major goal of the panels, our case is illustrative for an instance where the professions “*lay successful claim to normative dimensions of political processes*” (Townley, 1997: 280). They succeed in doing so by giving the impression of scrupulously taking into account all the principle’s criteria, while only a subset of them actually count in their decisions. So here again an agency problem grounds in the combination of a goal conflict and a monitoring problem (Eisenhardt, 1989).

The subsidy allocation process clearly comes to the fore as one that enacts two distinct agency relations: a principle is willing to support several agents to perform a task, the selection of whom is executed by another agent. Yet, what appears to be two agency dilemmas may in fact be a single bipartite treaty between policy makers and the art world.

Granting schemes (in their totality, so including criteria, but also panel composition, the periodicity of evaluations, etc.) can be comprehended as a specific aspect of organization praxis at the juncture of state and professional expert logics. We advance that in this artistic production environment, in the absence of a market’s logic, granting schemes and the interaction between professional evaluators and the state mediate the way in which a field is shaped. We have pointed at the discrepancy between the written evaluation reports and the factual decision of the expert panels that give the impression considering distinct policy goals, but are left to their own devices when it comes to the inclusion in and the exclusion from state sustenance. This may not be a deliberate tactic by which proficiency overrules politics, but as a tacit practice its implications for arts producers are not trivial. Diane Crane (1976) has advanced that those who control the key functions of the system set the cognitive and technical norms for innovation, on the basis of which they allocate rewards. Central to these resource allocation systems (“*reward systems*”) are norms for the production of new products. So in theory, in the absence of market selection, and in a production environment that is characterized by the scarcity of production means (as

is the case in the Netherlands, as argued by van Klink (2005), an argumentation that can be extended to the Flemish case) professionals dominating the resource allocation process have a major say about entry and exit, thus about the ecology of a field. This implies, in practice, that artistic developers are accountable to a small group of powerful professionals that have a strong impact on the performance of these producers. Thus, on the micro/organization level, being efficient in strategy or effective in resourcing allocation is far from a guarantee to survival. Organizations compete not just for customers and resources. Likewise conforming to societal rules or expectations may increase survival chances, since some degree of conformity ensures an organization's legitimacy within the external environment. Hence, from this environment stem formal and informal pressures exerted on organizations by other organizations upon which they are dependent and by cultural expectations within the society in which they operate (DiMaggio & Powell, 1991). These pressures may be especially palpable in the absence of market selection. Performance appraisals such as the one exposed, have been described as powerful institutional myths, part of the rationale of what organizations need to be legitimate (Townley, 1997: 266). Consequently the performance of non-profit art producers could be largely contingent on these institutionalized norms and may be moderated by organizational legitimacy. This in mind, the view advanced by Throsby and Withers (1979) in the wakening of cultural economics, of arts producing organizations as joint quality/quantity maximizers subject to a budget constraint, may be outmoded. Constraints extend beyond financial operations. If performance may be defined as the extent to which predetermined goals are achieved (Chow et al., 1994), the artistic performance of non-profits should be considered against the prevalent resource allocation systems. If these evaluation systems are inefficient, they may displace responsibilities and action, since organizations do not just compete for customers and resources, but also for political power and institutional legitimacy, and for social and economic fit (Aldrich, 1979; DiMaggio & Powell, 1991).

## References

- Aldrich, H. E. (1979). *Organizations and environments*. Englewood Cliffs, NJ: Prentice-Hall.  
Becker, H. S. (1974). Art as Collective Action. *American Sociological Review*, 39(6), 767-776.

- Bielby, D. D., Moloney, M., & Ngo, B. Q. (2005). Aesthetics of Television Criticism: Mapping Critics' Reviews in an Era of Industry Transformation. *Research in the Sociology of Organizations, 23*, 1-43.
- Bourdieu, P. (1983). The Field of Cultural Production, or the Economic World Reversed. *Poetics, 12*(4-5), 311-356.
- Bryman, A. (2008). *Social research methods* (third edition ed.). New York: Oxford University Press inc.
- Caves, R. E. (2000). *Creative Industries: Contracts between art and commerce*: Harvard University Press.
- Chow, G., Heaver, T. D., & Henriksson, L. E. (1994). Logistics Performance: Definition and Measurement. *International Journal of Physical Distribution & Logistics Management, Vol. 24* (No. 1), pp. 17-28.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3 ed.). New Jersey: Lawrence Erlbaum Associates.
- Deutsch, M. (1960). Trust, Trustworthiness, and the F Scale. *Journal of Abnormal and Social Psychology, 61*(1), 138-140.
- DiMaggio, P., & Hirsch, P. M. (1976). Production Organizations in Arts. *American Behavioral Scientist, 19*(6), 735-752.
- DiMaggio, P., & Powell, W. W. (1983). The Iron Cage Revisited - Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review, 48*(2), 147-160.
- DiMaggio, P., & Powell, W. W. (1991). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. In W. W. Powell & P. J. T. DiMaggio (Eds.), *The New Institutionalism in Organizational Analysis*. Chicago - London: The University of Chicago Press.
- Eisenhardt, K. M. (1989). Agency Theory - an Assessment and Review. *Academy of Management Review, 14*(1), 57-74.
- Feder, T., & Katz-Gerro, T. (2012). Who benefits from public funding of the performing arts? Comparing the art provision and the hegemony-distinction approaches. *Poetics, 40*(4), 359-381.
- Foster, P., Borgatti, S. P., & Jones, C. (2011). Gatekeeper search and selection strategies: Relational and network governance in a cultural market. *Poetics, 39*(4), 247-265.
- Friedland, R., & Alford, R. R. (1991). Bringing society back in: Symbols, practices and institutional contradictions. . In W. W. Powell & P. J. DiMaggio (Eds.), *The new institutionalism in organizational analysis* (pp. 232-263). Chicago, Il.: University of Chicago Press.
- Fullerton, D. (1991). On justifications for public support of the arts. *Journal of Cultural Economics, 15*(2), 67-82.
- Glynn, M. A., & Lounsbury, M. (2005). From the critics' corner: Logic blending, discursive change and authenticity in a cultural production system. *Journal of Management Studies, 42*(5), 1031-1055.
- Goodrick, E., & Reay, T. (2011). Constellations of Institutional Logics: Changes in the Professional Work of Pharmacists. *Work and Occupations, 38*(3), 372-416.
- Gorman, G. E. (2007). The Oppenheim effect in scholarly journal publishing. *Online Information Review, 31*(4), 417-419.

- Greenwood, R., Diaz, A. M., Li, S. X., & Lorente, J. C. (2010). The Multiplicity of Institutional Logics and the Heterogeneity of Organizational Responses. *Organization Science*, 21(2), 521-539.
- Greenwood, R., & Hinings, C. R. (1996). Understanding radical organizational change: Bringing together the old and the new institutionalism. *Academy of Management Review*, 21(4), 1022-1054.
- Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is originality in the humanities and the social sciences? *American Sociological Review*, 69(2), 190-212.
- Hadida, A. (2013 (forthcoming)). Performance in the Creative Industries. In C. Jones, M. Lorenzen & J. Sapsed (Eds.), *Oxford handbook of the Creative Industries* (pp. tba). Oxford: Oxford University Press.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* (Sixth Edition, International Edition ed.): Pearson Prentice Hall.
- Heinze, G. (2006). A comparative investigation of methods for logistic regression with separated or nearly separated data. *Statistics in Medicine*, 25(24), 4216-4226.
- Hirsch, P. M. (2000). Cultural industries revisited. *Organization Science*, 11(3), 356-361.
- Kassarjian, H. H. (1977). Content-Analysis in Consumer Research. *Journal of Consumer Research*, 4(1), 8-18.
- Kee, H. W., & Knox, R. E. (1970). Conceptual and Methodological Considerations in the Study of Trust and Suspicion. *The Journal of Conflict Resolution*, 14(3), 357-366.
- Kolbe, R. H., & Burnett, M. S. (1991). Content-Analysis Research - an Examination of Applications with Directives for Improving Research Reliability and Objectivity. *Journal of Consumer Research*, 18(2), 243-250.
- Krebs, S., & Pommerehne, W. W. (1995). Politico-economic interactions of German public performing arts institutions. *Journal of Cultural Economics*, 19(1), 17-32.
- Krippendorff, K. (1980). *Content Analysis. An Introduction to Its Methodology*. California: SAGE Publications, Inc.
- Krippendorff, K., & Bock, M. A. (2009). *The content analysis reader*. Thousand Oaks, California: SAGE Publications, Inc.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models* (international edition, fifth edition ed.). New York: McGraw-Hill/Irwin.
- Levin-Rozalis, M. (2000). Abduction: A Logical Criterion For programme and Project Evaluation. *Evaluation*, 6(4), 415-432.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An Integrative Model of Organizational Trust. *Academy of Management Review*, 20(3), 709-734.
- Menard, S. (2002). *Applied logistic Regression Analysis* (second edition ed.). Thousand Oaks, CA: Sage.
- Oakes, L. S., Townley, B., & Cooper, D. J. (1998). Business planning as pedagogy: Language and control in a changing institutional field. *Administrative Science Quarterly*, 43(2), 257-292.
- OCW/Boekmanstudies, M. v. (2007). Cultuurbeleid in Nederland. In C. a. S. Education (Ed.) (pp. 192). Den haag/Amsterdam.
- Peterson, R. A., & Anand, N. (2004). The production of culture perspective. *Annual Review of Sociology*, 30, 311-334.
- Reay, T., & Hinings, C. R. (2005). The recomposition of an organizational field: Health care in Alberta. *Organization Studies*, 26(3), 351-384.
- Reay, T., & Hinings, C. R. (2009). Managing the Rivalry of Competing Institutional Logics. *Organization Studies*, 30(6), 629-652.

- Riffe, D., Lacy, S., & Fico, F. G. (2005). *Analyzing Media Messages. Using Quantitative Content Analysis in Research*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Schurz, G. (2008). Patterns of abduction. *Synthese*, 164(2), 201-234.
- Scott, W. R. (1987). The Adolescence of Institutional Theory. *Administrative Science Quarterly*, 32(4), 493-511.
- Seifert, M., & Hadida, A. L. (2006). Facilitating talent selection decisions in the music industry. *Management Decision*, 44(6), 790-808.
- Shrum, W. (1991). Critics and Publics - Cultural Mediation in Highbrow and Popular Performing Arts. *American Journal of Sociology*, 97(2), 347-375.
- Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education*, 46(1), 29-48.
- Suchman, M. C. (1995). Managing Legitimacy - Strategic and Institutional Approaches. *Academy of Management Review*, 20(3), 571-610.
- Throsby, D. (1982). Economics and the Arts - a Review of 7 Years. *Economic Record*, 58(162), 242-252.
- Throsby, D. (1994). The Production and Consumption of the Arts - a View of Cultural Economics. *Journal of Economic Literature*, 32(1), 1-29.
- Throsby, D., & Withers, G. (1979). *The economics of the performing arts*. New York: St Martin's Press.
- Townley, B. (1997). The institutional logic of performance appraisal. *Organization Studies*, 18(2), 261-285.
- Turbide, J., & Laurin, C. (2009). Performance Measurement in the Arts Sector: the Case of the Performing Arts. *International Journal of Arts Management*, 11(2), 56-70.
- van Klink, P. (2005). *Kunsteconomie in nieuw perspectief. Rijkskunstbeleid beoordeeld*. Groningen: K's Concern.
- Venkatraman, N., & Ramanujam, V. (1986). Measurement of Business Performance in Strategy Research - a Comparison of Approaches. *Academy of Management Review*, 11(4), 801-814.
- Vittinghoff, E., & McCulloch, C. E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. *American Journal of Epidemiology*, 165(6), 710-718.
- Voss, G. B., Cable, D. M., & Voss, Z. G. (2000). Linking organizational values to relationships with external constituents: A study of nonprofit professional theatres. *Organization Science*, 11(3), 330-347.
- Werck, K., & Heyndels, B. (2007). Programmatic choices and the demand for theatre: the case of Flemish theatres. *Journal of Cultural Economics*, 31(1), 25-41.
- White, H. C., & White, C. A. (1993 (1965)). *Canvases and Careers: Institutional Change in the French Painting World*. Chicago: University of Chicago Press.
- Wijnberg, N. M. (2011). Classification systems and selection systems: The risks of radical innovation and category spanning. *Scandinavian Journal of Management*, 27(3), 297-306.
- Zan, L., Bonini Baraldia, S., Ferria, P., Lusiania, M., & Mariania, M. M. (2012). Behind the scenes of public funding for performing arts in Italy: hidden phenomena beyond the rhetoric of legislation. *International Journal of Cultural Policy* 18(1), 76-92.
- Zand, D. E. (1972). Trust and Managerial Problem Solving. *Administrative Science Quarterly*, 17(2), 229-239.

Criteria	Belgium (Flanders)	The Netherlands	PERFORMANCE DIMENSIONS
ARTISTIC QUALITY	Creations that are intriguing, innovative, strong, emotional, beautiful, relevant, original; honest and authentic theatre; important, inventive, talented, devoted creators; a theatre language that is expressive, unique, original, important, strong; devotion to the craft of theatre; the dramaturgy is strong/poor; texts are original, made accessible, communicative... (3)	craftsmanship, skills of team members, originality, artistic signature, vision of the artistic leader, distinguished, expressive power, challenge imagination, the creation of meaningful performances (1)	Quality and reputation
PROFILING & POSITION	sharp/strong profile, unique position, special, coherent, (no longer) relevant or unique, remarkable, original, obstinate, having played a (pioneering) role, having been part of..., one of the representatives of..., distinguishing... (1)	different, unique (2)	
(INTER)NATIONAL REPUTATION	Spreading, dispersal, dissemination, international operations, abroad, (limited) visibility, presence/absence (in specific countries, in national theatres...), touring (within a country, around national theatres...), travelling to, (inter)national reputation, increase of attraction, disposing of a network of European partners, mapping internationally (also 1), crossing regional/national borders (4)	outreach, dispersal, distribution (3)	
REGIONAL ADDED VALUE	regional meaning/significance, regional recognition, radiance, embeddedness, anchoring (10)	link to contribution to the distribution (3)	Social value and market
AUDIENCE OUTREACH	target groups/audience, arts education, public relations, audience recruitment/growth (7)	Audience development, education, marketing (5)	
DIVERSITY AND INTERCULTURALITY	New Belgians, multicolored and multilingual environment, non-western cultures, diverse backgrounds (13)	cultural diversity (intercultural supply and program) (2)	
SOCIAL VALUE	social-artistic activities/dynamics, social importance, socially critical and political themes/language/voice, social engagement (11)	-	
COLLABORATION AND NETWORKING	collaborations, network, alliances, co-productions, synergies (5)	-	Efficiency
CHAIN IDEA	-	the creation, production, programming and the audience development, exemplified by collaborations (4)	
CULTURAL ENTREPRENEURSHIP & MANAGEMENT	-	market income, playlist, attitude (5)	
LONG-TERM VISION	plans, (lack of) vision, policy (2)	-	Strategy and planning
REALISM	we do not find explicit synonyms, but do find combinations of words that reflect this instance (9)	-	
FEASIBILITY	the plan is (not) feasible (6)	-	
() between brackets is the number of the original criterion			

Table 1: evaluation criteria for theatre organizations in Flanders and the Netherlands

Criteria	dimensions	Belgium (n=57)		freq % (rank)	The Netherlands (n=84)		freq % (rank)
		Presence N			Presence N		
<b>Artistic quality</b>		<b>57</b>	<b>100%</b>		<b>84</b>	<b>100%</b>	
<b>Profile and position</b>	profile or position	<b>47</b>	<b>81%</b>	<b>5</b>	<b>84</b>	<b>100%</b>	<b>1</b>
	profile	29	50%		17	20%	
	position	38	66%		25	30%	
	unique	21	36%		12	14%	
	distinctive				16	19%	
	complementary				26	31%	
	meaningful				4	5%	
	significant	30	52%				
	location	10	17%			0%	
	important	6	10%		8	10%	
	strong	5	9%		1	1%	
	remarkable	1	2%			0%	
	key player	3	5%		3	4%	
	recognizable				3	4%	
	visible	6	10%			0%	
	exemplary				2	2%	
	pioneering (first)	3	5%		5	6%	
	innovative				2	2%	
	mature				1	1%	
<b>(inter)national reputation</b>		<b>53</b>	<b>91%</b>	<b>1</b>	<b>36</b>	<b>43%</b>	<b>6</b>
	national	40	69%		26	31%	
	international	40	69%		22	26%	
<b>regional added value</b>		<b>28</b>	<b>48%</b>	<b>8</b>			
	region	16	28%				
	City	12	21%				
<b>audience range</b>		<b>49</b>	<b>84%</b>	<b>4</b>	<b>54</b>	<b>64%</b>	<b>2</b>
	children	11	19%		9	11%	
	youth	18	31%		15	18%	
	traditionally high	4	7%		2	2%	
	traditionally low	12	21%		3	4%	
	new	11	19%		21	25%	
	social groups	19	33%		4	5%	
	divers	29	50%		9	11%	
	large				22	26%	
<b>interculturality &amp; diversity</b>		<b>40</b>	<b>69%</b>	<b>6</b>	<b>22</b>	<b>26%</b>	<b>9</b>
<b>social value</b>		<b>38</b>	<b>66%</b>	<b>7</b>	<b>24</b>	<b>29%</b>	<b>8</b>
	content	22	38%		18	21%	
	outreach	23	40%		10	12%	
<b>collaboration and networking</b>		<b>52</b>	<b>90%</b>	<b>2</b>	<b>51</b>	<b>61%</b>	<b>4</b>
	to economize	16	28%				
	to create staging opportunities	41	71%				
	to develop (coproductions)	45	78%				
<b>chain idea</b>					<b>54</b>	<b>64%</b>	<b>3(2)</b>
<b>cultural entrepreneurship and business</b>					<b>42</b>	<b>50%</b>	<b>5</b>
<b>long term vision</b>	vision	<b>22</b>	<b>38%</b>	<b>9</b>	<b>26</b>	<b>31%</b>	<b>7</b>
	plan	35	60%		18	21%	
<b>realism</b>		<b>11</b>	<b>19%</b>	<b>10</b>	<b>6</b>	<b>7%</b>	<b>10</b>
<b>feasibility</b>		<b>50</b>	<b>86%</b>	<b>3</b>			

Table 2: the occurrence of the criteria in the evaluations in Flanders and the Netherlands

Table 3: correlation table of the criteria in the evaluations in Flanders

**Means, standard deviations and Spearman's Rho Correlations among all variables**

Variables	Mean	S.d.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<i>independent variables</i>																
1 Artistic quality	3.31	1.158	1.000													
2 Position	.66	.479	.572**	1.000												
3 Reputation	.74	.442	.616**	.566**	1.000											
4 Audience	.84	.365	.125	.090	.291*	1.000										
5 Diversity	.52	.504	.155	.243	.217	.062	1.000									
6 Vision	.31	.467	.252	.330*	.311*	.082	.201	1.000								
7 Feasibility	.69	.467	.528**	.533**	.455**	.021	.396**	.208	1.000							
8 Realism	.19	.395	.206	.166	.185	.086	-.061	.056	.039	1.000						
9 Collaboration	.69	.467	-.211	-.173	-.056	.021	.023	-.033	-.208	-.341**	1.000					
10 Regional value	.47	.503	-.008	.023	-.238	-.173	.141	.121	.178	-.099	-.046	1.000				
11 Social	.57	.500	.021	.174	.122	.204	.274*	-.093	.093	-.023	-.057	.114	1.000			
<i>dependent variables</i>																
12 Judgment (before response)	3.53	1.341	.779**	.662**	.543**	.144	.307*	.268*	.652**	.250	-.290*	.089	.109	1.000		
13 Judgment (after response)	3.67	1.316	.778**	.679**	.531**	.115	.232	.295*	.576**	.331*	-.281*	.046	.106	.970**	1.000	
14 Judgment (after, dummy)	.78	.421	.644**	.741**	.438**	.112	.308*	.361**	.622**	.155	-.271*	.170	.200	.735**	.756**	1.000

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

12 & 13: after receiving the evaluation outcome, organizations are allowed to react, after which panels may reconsider their decisions. This happens occasionally.

Table 4: correlation table of the criteria in the evaluations in the Netherlands

**Means, standard deviations and Spearman's Rho Correlations among all variables**

<i>Variables</i>	Mean	S.d.	1	2	3	4	5	6	7	8
<i>independent variables</i>										
1 Artistic quality	2.51	.844	1.000							
2 Position	.33	.473	.528**	1.000						
3 Reputation	.43	.498	.611**	.340**	1.000					
4 Audience	.66	.477	.356**	.121	.309**	1.000				
5 Diversity	.27	.446	.341**	.220*	.257*	.262*	1.000			
6 Chain idea	.63	.485	.102	.209	.195	.094	.117	1.000		
7 Cultural entrepreneurship	.35	.481	.406**	.350**	.290**	.264*	-.045	.244*	1.000	
<i>dependent variable</i>										
8 Judgment	.50	.503	.887**	.597**	.715**	.463**	.385**	.253*	.485**	1.000

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

Figure 1: scatter plot artistic quality against judgment (Dutch data)

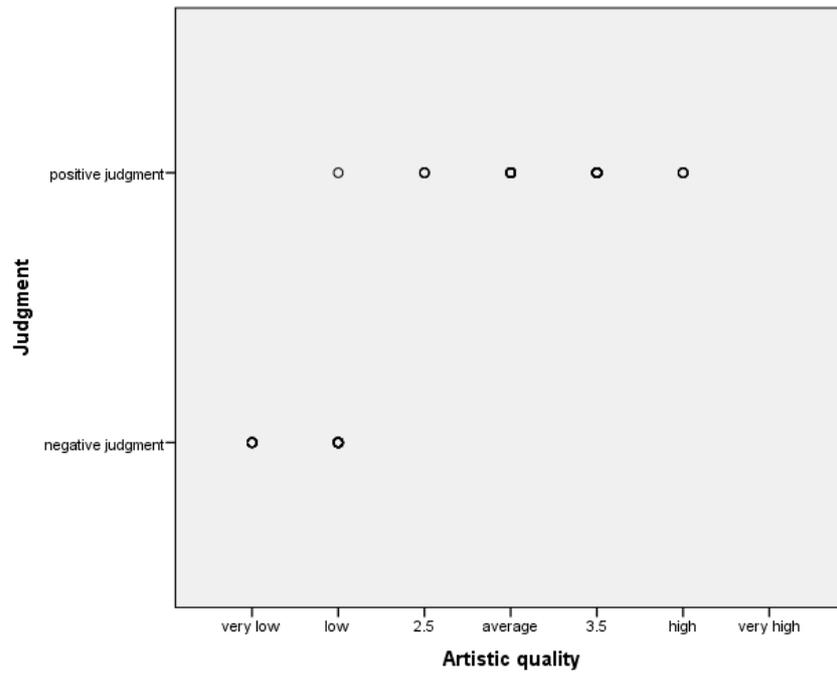


Figure 2: scatter plot artistic quality against judgment (Flanders data)

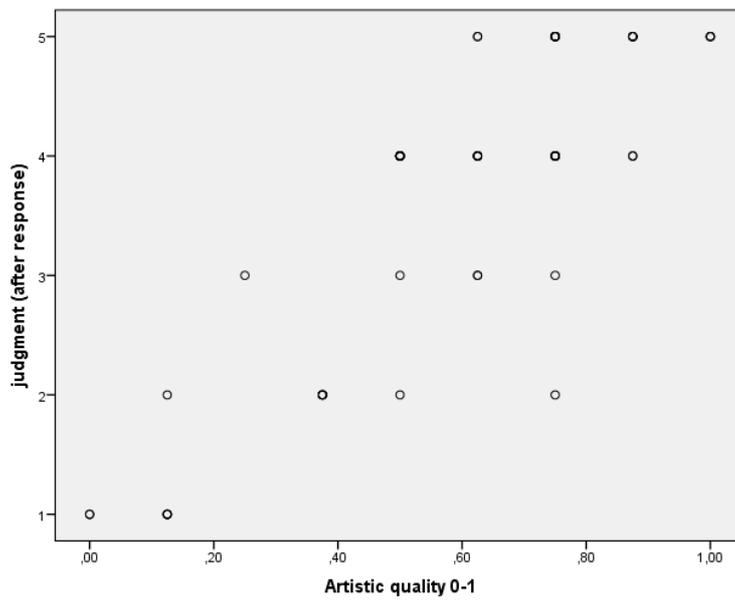


Table 5: regression table (the Netherlands)

	M1	M2
Criteria		
Constant	-7.920***	-11.970***
Artistic Quality	8.740***	24.779***
Position	1.875	
Reputation	2.393*	3.197*
Chain Idea	1.183	
Audience Orientation	1.780	
Diversity	0.435	
Cultural Entrepreneurship	-0.612	
Overall model evaluation		
Likelihood ratio test	90.749*** df7	97.632*** df2
Wald test	22.091** df7	7.842** df2

\*=p<.05; \*\*=p<.01; \*\*\*=p<.001

Table 6: regression table (Flanders)

	M1	M2
Criteria	Logit	Logit
Threshold (judgment = 1)	-3.963	-1.695
Threshold (judgment = 2)	0.618	1.588
Threshold (judgment = 3)	2.319	3.03
Threshold (judgment = 4)	6.149	6.852
Artistic Quality	11.669***	10.680***
Position	-2.373**	-2.406**
Reputation	1.166	
Regional Importance	-0.485	
Audience Orientation	-0.327	
Diversity	-0.367	
Social Realm	-0.558	
Realism	-1.796*	-1.390
Feasibility	-0.918	
Vision	-0.797	
Collaboration	0.313	
Overall model evaluation		
Chi <sup>2</sup>	83.561***	76.600***
Goodness-of-fit test		
Pearson	108.326 df 209 (P=1.000)	51.6880 df 65 (p=0.885)
pseudo R <sup>2</sup>		
Cox and Snell	0.7630	0.733
Nagelkerke	0.8090	0.777
McFadden	0.5030	0.461

\*=p<.05; \*\*=p<.01; \*\*\*=p<.001